

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 8, 2003, 20:55:17 ; Search time 2937 Seconds
(without alignments)

15879.120 Million cell updates/sec

Title: US-10-007-527A-1

Sequence: 1 atgaccagcgtgaagtgcctga.....taacatcggtcgcgaagtga 1140

Scoring table:

OLIGO_NUC

Gapop 60.0 , Gapext 60.0

Word size : 0

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: listing first 45 summaries

Database : GenBank:

1: gb_ba:*
2: gb_ncg:*
3: gb_in:*
4: gb_cm:*
5: gb_ov:*
6: gb_pat:*
7: gb_pn:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vl:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rnd:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1140	100.0	1140	6	AX548642 Sequence
2	1140	100.0	6334	6	AY178757 Rhodococc
3	1140	100.0	6334	6	AX548646 Sequence
4	1140	100.0	9652	6	AX548648 Sequence
5	1140	100.0	11241	6	AX548647 Sequence
6	1140	100.0	11241	12	AY180162 Shuttle v
7	22	1.9	202897	2	AC133415 Rattus no
8	22	1.9	224046	2	AC123219 Rattus no
9	22	1.9	228864	2	AC103497 Rattus no
10	21	1.8	147405	2	AP005161 Rattus no
11	21	1.8	155547	2	AP005799 Rattus no
12	20	1.8	167	6	AX548660 Sequence
13	20	1.8	167	6	AX049359 Sequence
14	20	1.8	1402	9	AY227111 Homo sapi
15	20	1.8	1420	9	AY227112 Homo sapi
16	20	1.8	1428	6	AX049362 Sequence
17	20	1.8	1501	6	AX717763 Sequence
18	20	1.8	1513	6	AX717761 Sequence
19	20	1.8	1961	6	AX714753 Sequence
20	20	1.8	1961	6	AK057340 Homo sapi
21	20	1.8	3114	6	AX686980 Sequence
22	20	1.8	5585	1	AF516142 Homo sapi
23	20	1.8	10888	1	AE009218 Agrobacte
24	20	1.8	11693	1	AE008185 Agrobacte
25	20	1.8	89319	9	HSBRK29F11 Human DNA
26	20	1.8	94625	2	AC011337 Homo sapi
27	20	1.8	121727	2	AC119124 Rattus no
28	20	1.8	133614	2	AC025433 Homo sapi
29	20	1.8	154520	10	AC132098 Mus muscu
30	20	1.8	159562	9	AC034205 Homo sapi
31	20	1.8	181772	2	AC007689 Homo sapi
32	20	1.8	201659	2	AC120558 Mus muscu
33	20	1.8	208523	2	AC113532 Mus muscu
34	20	1.8	241150	1	AC105884 Rattus no
35	20	1.8	305520	1	AE016780 Pseudomon
36	19	1.7	642	8	AY257541 Brascica
37	19	1.7	882	8	SAU25696 Sinapis alb
38	19	1.7	891	4	OCABP1 Rabbit mRNA
39	19	1.7	1092	9	HSHBRG1 Human mRNA
40	19	1.7	1192	9	HSHBRG2 Human mRNA
41	19	1.7	1246	4	OCU57553 Oryctolagus
42	19	1.7	3390	4	AF144711 Oryctolagus
43	19	1.7	3810	6	AX411065 Sequence
44	19	1.7	3810	9	HSHBRG Human gene
45	19	1.7	4579	14	SCSCCATMNA Saccharomyc

ALIGNMENTS

RESULT 1
LOCUS AX548642 1140 bp DNA linear PAT 27-NOV-2002
DEFINITION Sequence 1 from Patent WO02055709.
ACCESSION AX548642
VERSION AX548642.1 GI:25813612
KEYWORDS
SOURCE
ORGANISM
Rhodococcus erythropolis
Rhodococcus erythropolis
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Corynebacterineae; Nocardiaceae; Rhodococcus.

REFERENCE
AUTHORS
TITLE
JOURNAL
Bramucci, M.G., Cheng, Q., Kostichka, K.N. and Tomb, J.F.
Rhodococcus cloning and expression vectors
Patent: NO 02055709-A 1 18-JUL-2002

Pred. No. is the number of results predicted by chance to have a

FEATURES
source

E.I. DU PONT DE NEMOURS AND COMPANY (US)
Location/Qualifiers
1. 1140
/organism="Rhodococcus erythropolis"
/mol_type="genomic DNA"
/strain="PAN12"
/db_xref="taxon:1833"

BASE COUNT 241 a 287 c 347 g 265 t

ORIGIN

Query Match 100.0%; Score 1140; DB 6; Length 1140;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1140; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

Qy 1 ATGACCGAGCTGAGTCTGTAACACCTTCCGCGCAAAACCGGCTCCCGTCTCGTCTG 60
Db 1 ATGACCGAGCTGAGTCTGTAACACCTTCCGCGCAAAACCGGCTCCCGTCTCGTCTG 60

Qy 61 TCCGATTAAGCGCGGATCCGCGACGAACTGCGACCCAACTTCAACAAATCACCAGTCA 120
Db 61 TCCGATTAAGCGCGGATCCGCGACGAACTGCGACCCAACTTCAACAAATCACCAGTCA 120

Qy 121 GAAACATTTAACGCTGTGCGCGCGCGAATTTCTGCGCTGAAACGCTGTGACCAATTGTCAC 180
Db 121 GAAACATTTAACGCTGTGCGCGCGCGAATTTCTGCGCTGAAACGCTGTGACCAATTGTCAC 180

Qy 181 GGTCCGAAAGGTTCTGAAATTCGAGGCGCTTCTGCGGAAAGGCGTGAATCTGCCCC 240
Db 181 GGTCCGAAAGGTTCTGAAATTCGAGGCGCTTCTGCGGAAAGGCGTGAATCTGCCCC 240

Qy 241 TGCCTGCGGAAAGTCTGAGTGCACATCTGACAGCAAAATTTCTCAAGTTGTTGCTCAT 300
Db 241 TGCCTGCGGAAAGTCTGAGTGCACATCTGACAGCAAAATTTCTCAAGTTGTTGCTCAT 300

Qy 301 CAACCTCGGACCTGAGATCTGTTCGATGATGACGATGACCAATGCGCATACAGCTGTGAC 360
Db 301 CAACCTCGGACCTGAGATCTGTTCGATGATGACGATGACCAATGCGCATACAGCTGTGAC 360

Qy 361 CGGCTCCACGACCTTATGAGTCTGAGCTTTCCGCGACCTGCGAAAGTGTGACCAACGGTCT 420
Db 361 CGGCTCCACGACCTTATGAGTCTGAGCTTTCCGCGACCTGCGAAAGTGTGACCAACGGTCT 420

Qy 421 CGTTGCGGTACGGAACCTGGAATGTAACGCTGCGACGATGCTGCGCGCTGTTGAATC 480
Db 421 CGTTGCGGTACGGAACCTGGAATGTAACGCTGCGACGATGCTGCGCGCTGTTGAATC 480

Qy 481 ACTACCGAAAAAAGCGCTGCGACGTCACGCTTACCGCGTACTCATGTTCAATGTTGAC 540
Db 481 ACTACCGAAAAAAGCGCTGCGACGTCACGCTTACCGCGTACTCATGTTCAATGTTGAC 540

Qy 541 GTGAGTGAAGAACATCTCTGAAATCTTCTCGGATGCGATGTTCCATCGGTGACCTTCCAAA 600
Db 541 GTGAGTGAAGAACATCTCTGAAATCTTCTCGGATGCGATGTTCCATCGGTGACCTTCCAAA 600

Qy 601 CTGCTATCTCTGGGATTTGCTGCGGCACTAGCTAATTCGAGTGTGCTCGATGACGAAAG 660
Db 601 CTGCTATCTCTGGGATTTGCTGCGGCACTAGCTAATTCGAGTGTGCTCGATGACGAAAG 660

Qy 661 ATCGCGGTGAGTGTGATCAAGTTCTGCTGCGATCTGACGAAATTCATCTGCGCTT 720
Db 661 ATCGCGGTGAGTGTGATCAAGTTCTGCTGCGATCTGACGAAATTCATCTGCGCTT 720

Qy 721 GGTATGAGGTTGTTGATGAGGCGAAGGAAATGTTGTCACATGTCACACCGTGG 780
Db 721 GGTATGAGGTTGTTGATGAGGCGAAGGAAATGTTGTCACATGTCACACCGTGG 780

Qy 781 GAAATCGCTGTTGATGAGGCGGAGATCCACAAGGTTGAACTGTGCGGAAATTT 840
Db 781 GAAATCGCTGTTGATGAGGCGGAGATCCACAAGGTTGAACTGTGCGGAAATTT 840

Qy 841 GAGTTTGGTTGATGAGGCGTCCGCGCAATTCGCTGTGATTCGCTGCGGAGCT 900
Db 841 GAGTTTGGTTGATGAGGCGTCCGCGCAATTCGCTGTGATTCGCTGCGGAGCT 900

```

Qy 901 GGTCTTTGGGCGAACTTAACAGATGCTAGATCTGTTAGCAGAAAGATCTCCCGGCTC 960
Db 901 GGTCTTTGGGCGAACTTAACAGATGCTAGATCTGTTAGCAGAAAGATCTCCCGGCTC 960

Qy 961 ATGTTGCGATCATTTCCGCGCGGATCGTGAATGATGATTCGACCTTGTGCGCTTACGTC 1020
Db 961 ATGTTGCGATCATTTCCGCGCGGATCGTGAATGATGATTCGACCTTGTGCGCTTACGTC 1020

Qy 1021 TTGCGGAGATCTCGGATCTCGTGAAGTGTGCGGCGACCTTGGGAAATCTTGTGATCAC 1080
Db 1021 TTGCGGAGATCTCGGATCTCGTGAAGTGTGCGGCGACCTTGGGAAATCTTGTGATCAC 1080

Qy 1081 TTGCATTATCGATTCGCGCGAGGATGTGCGGCGCCCGCATTAATTCGTTGCGCAAGTGA 1140
Db 1081 TTGCATTATCGATTCGCGCGAGGATGTGCGGCGCCCGCATTAATTCGTTGCGCAAGTGA 1140

RESULT 2
AY178757/c 6334 bp DNA linear BCT 01-MAR-2003

LOCUS
DEFINITION Rhodococcus erythropolis cryptic plasmid PAN12, complete sequence.
ACCESSION AY178757
VERSION AY178757.1 GI:28628256
KEYWORDS
SOURCE Rhodococcus erythropolis
ORGANISM Rhodococcus erythropolis
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales; Corynebacterineae; Nocardiaceae; Rhodococcus.
Koslichka, K., Tao, L., Brammucci, M., Tomb, J.-F., Nagarajan, V. and Cheng, Q.
A small cryptic plasmid from Rhodococcus erythropolis: Characterization and utility for gene expression
Unpublished
2 (bases 1 to 6334)
Koslichka, K., Tao, L., Brammucci, M., Tomb, J.-F., Nagarajan, V. and Cheng, Q.
Direct Submission
Submitted (11-NOV-2002) CR&D, E.I. DuPont de Nemours Inc., Experimental Station, Wilmington, DE 19880-0328, USA

TITLE
AUTHORS
REFERENCE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
FEATURES
source

1. 6334
/organism="Rhodococcus erythropolis"
/mol_type="genomic DNA"
/db_xref="taxon:1833"
/plasmid="PAN12"
/note="cryptic plasmid"
/gene="rep"
complement (1912..3051)
/gene="rep"
complement (1912..3051)
/note="Rep"
/codon_start=1
/transl_table=11
/product="repllication protein"
/protein_id="AA031602.1"
/db_xref="GI:28628257"

translation="MTSVSAEHLGSKDRPEVLYSSDKGTRHRLRPEKQITTSRTN
ACGRPIISGVNAGVTIVNGPKSGFGGLRSCKGWI CPCAGKVAHRAHRADEISQVVAHQL
GTGSVAMTWMTMHTAGORLHDHTGLISAAMKAATNGRRTREMYGCDGYRAVABI
THGKNMGVHVHVALMLMPSGDSNLTLESFSDAMPDRTSLVLSGFAPAPRNSGGLDV
RKIGSEADQVLAAYLTRIASGVMEVSGSGSGRGRNRAPEIAYDAVGDQOALEL
WRPEFGSMKRAIAMSRLGRABAGAEITDQIYVQESRAVYMAAILPARSMWIR
TCAPYFGBELGIVGAGTMENTRDLHTPLPADVAPPIISVRK"

gene
CDS
/note="div"
/gene="div"
complement (4288..5178)
/gene="div"
/note="div"
/codon_start=1
/transl_table=11
/product="putative cell division protein"

```
/protein_id="AA031603.1"
/db_xref="GI:28628258"
/translation="MDQTDPIPIAIGMNELAQPVWDIAKDAHMLIOGKTRSGKSQC
TYMLLAQAGSNPAIVYGVDPISVLLAFPHRPAEPNLEIANDPDKYLRFQFPA
ESDRLECFMDRIDIKISLESPALPLILAEFPGLIEGAOFDNTNGKPADRYAP
RTSLVROIAAOSAKAGIMILLIAQRAEASIVGASNRSAVMTYRVPBPSVRLH
PNATPEBCALVEFPVPGQGFDPQGRKQRTVRVGEISTIASIYENADLATEALIN
IDRAQMTASEPPLHLDIG"
BATCH COUNT 1386 a 1807 c 1681 g 1460 t
ORIGIN

Query Match 100.0%; Score 1140; DB 1; Length 6334;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1140; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGACGAGCGTAAGTGTCTGAACAACCTTCCGCGAAGACCGGCTCCCGTCTGATGTCG 60
Db 3051 ATGACGAGCGTAAGTGTCTGAACAACCTTCCGCGAAGACCGGCTCCCGTCTGATGTCG 2992

QY 61 TCCGATTAAGCGCGCATCCGGCAAGACCTGCGAACCCAACTTCAACAAATCACACGTCA 120
Db 2991 TCCGATTAAGCGCGCATCCGGCAAGACCTGCGAACCCAACTTCAACAAATCACACGTCA 2932

QY 121 GAAACATTTAAACCCCTGTGCGCCGCGCATTTCTGCGGTGAACGCTGACCATTTGTCAAC 180
Db 2931 GAAACATTTAAACCCCTGTGCGCCGCGCATTTCTGCGGTGAACGCTGACCATTTGTCAAC 2872

QY 181 GGTCCGAAAGGTTCTGGATTCTGAGAGGCTTCTGTTCTCGCGAAAGGGCTGGATCTGCCC 240
Db 2871 GGTCCGAAAGGTTCTGGATTCTGAGAGGCTTCTGTTCTCGCGAAAGGGCTGGATCTGCCC 2812

QY 241 TGTCTGTCGGGAAAAGTCGTGTGACATCTGACAGACGAAATTTCTCAAGTTGTGCTCAT 300
Db 2811 TGTCTGTCGGGAAAAGTCGTGTGACATCTGACAGACGAAATTTCTCAAGTTGTGCTCAT 2752

QY 301 CAACCTGGGACTGGATCTGTTGGATGTGACATGACATCCATGCGCCATACAGCTGGTCAG 360
Db 2751 CAACCTGGGACTGGATCTGTTGGATGTGACATGACATCCATGCGCCATACAGCTGGTCAG 2692

QY 361 CGGCTCCAGACCTATGACATGACATTTGGGAGCGCTGGAAGCTGCGACCAACGCGTCT 420
Db 2691 CGGCTCCAGACCTATGACATGACATTTGGGAGCGCTGGAAGCTGCGACCAACGCGTCT 2632

QY 421 CGTTGGCGTACGGAACCTGAATGTACGCGTGCACGGAATACGTGCGCGCTGTGAAATC 480
Db 2631 CGTTGGCGTACGGAACCTGAATGTACGCGTGCACGGAATACGTGCGCGCTGTGAAATC 2572

QY 481 ACTACGGAAGAAAACGCGTGCACGTTCACGCGTCACTCATGTTCAAGTGTGAC 540
Db 2571 ACTACGGAAGAAAACGCGTGCACGTTCACGCGTCACTCATGTTCAAGTGTGAC 2512

QY 541 GTGAGTGAAGAACTCCCGAATCCCTCGAATGCGATGCGATGTTCAATCGGTGGAACCTTCCAAA 600
Db 2511 GTGAGTGAAGAACTCCCGAATCCCTCGAATGCGATGCGATGTTCAATCGGTGGAACCTTCCAAA 2452

QY 601 CTGTAATCTCTGGGATTTGCTGCGCACCTACGTAATCGGGTGGTCTCGATGTAAGAAAG 660
Db 2451 CTGTAATCTCTGGGATTTGCTGCGCACCTACGTAATCGGGTGGTCTCGATGTAAGAAAG 2392

QY 661 ATGGCGGCTGAAGCTGATCAAGTTCTGCTGCGTATCTGAAGAAATTTGATCTGGCGT 720
Db 2391 ATGGCGGCTGAAGCTGATCAAGTTCTGCTGCGTATCTGAAGAAATTTGATCTGGCGT 2332

QY 721 GGTATGAGAGTTGTGATGCGACGAGAAAAGTGTGACATGCGAACCGTGCACCTGCG 780
Db 2331 GGTATGAGAGTTGTGATGCGACGAGAAAAGTGTGACATGCGAACCGTGCACCTGCG 2272

QY 781 GAAATCCTGTTGATGAGTGGCGGGATTCACAACGTTGAACTGTGAGCGAGAAATTT 840
Db 2271 GAAATCCTGTTGATGAGTGGCGGGATTCACAACGTTGAACTGTGAGCGAGAAATTT 2212

QY 841 GAGTTTGTTGATGAGGAGCTCGGCAATGCGGTGCTCCGTGGAATTTGCTGCCAGCT 900
Db 841 GAGTTTGTTGATGAGGAGCTCGGCAATGCGGTGCTCCGTGGAATTTGCTGCCAGCT 900

RESULT 3
AX548646/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLES
JOURNAL
FEATURES
source
1. .6334
/organism="Rhodococcus erythropolis"
/mol_type="genomic DNA"
/strain="AN12"
/db_xref="taxon:1833"
BATCH COUNT 1386 a 1807 c 1681 g 1460 t
ORIGIN

Query Match 100.0%; Score 1140; DB 6; Length 6334;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1140; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGACGAGCGTAAGTGTCTGAACAACCTTCCGCGAAGACCGGCTCCCGTCTGATGTCG 60
Db 3051 ATGACGAGCGTAAGTGTCTGAACAACCTTCCGCGAAGACCGGCTCCCGTCTGATGTCG 2992

QY 61 TCCGATTAAGCGCGCATCCGGCAAGACCTGCGAACCCAACTTCAACAAATCACACGTCA 120
Db 2991 TCCGATTAAGCGCGCATCCGGCAAGACCTGCGAACCCAACTTCAACAAATCACACGTCA 2932

QY 121 GAAACATTTAAACCCCTGTGCGCCGCGCATTTCTGCGGTGAACGCTGACCATTTGTCAAC 180
Db 2931 GAAACATTTAAACCCCTGTGCGCCGCGCATTTCTGCGGTGAACGCTGACCATTTGTCAAC 2872

QY 181 GGTCCGAAAGGTTCTGGATTCTGAGAGGCTTCTGTTCTCGCGAAAGGGCTGGATCTGCCC 240
Db 2871 GGTCCGAAAGGTTCTGGATTCTGAGAGGCTTCTGTTCTCGCGAAAGGGCTGGATCTGCCC 2812

QY 241 TGTCTGTCGGGAAAAGTCGTGTGACATCTGACAGACGAAATTTCTCAAGTTGTGCTCAT 300
Db 2811 TGTCTGTCGGGAAAAGTCGTGTGACATCTGACAGACGAAATTTCTCAAGTTGTGCTCAT 2752

QY 301 CAACCTGGGACTGGATCTGTTGGATGTGACATGACATCCATGCGCCATACAGCTGGTCAG 360
Db 2751 CAACCTGGGACTGGATCTGTTGGATGTGACATGACATCCATGCGCCATACAGCTGGTCAG 2692

QY 361 CGGCTCCAGACCTATGACATGACATTTGGGAGCGCTGGAAGCTGCGACCAACGCGTCT 420
Db 361 CGGCTCCAGACCTATGACATGACATTTGGGAGCGCTGGAAGCTGCGACCAACGCGTCT 420
```


QY 1021 TTGGCGGAGATCTTGGAGCTGTCGAGAGTGGCGGAGCTTGGGAAAATCTTCGTATAC 1080
Db 4032 TTGGCGGAGATCTTGGAGCTGTCGAGAGTGGCGGAGCTTGGGAAAATCTTCGTATAC 3973
QY 1081 TTGCATTATGATGATGCGGAGCGGAGTGGCGGAGCGGAGTAAATATGATGCGGAGTGA 1140
Db 3972 TTGCATTATGATGATGCGGAGCGGAGTGGCGGAGCGGAGTAAATATGATGCGGAGTGA 3913

RESULT 5
AX548647/c

LOCUS AX548647 11241 bp DNA linear PAT 27-NOV-2002
DEFINITION Sequence 6 from Patent WO02055709.
ACCESSION AX548647
VERSION AX548647.1 GI:25813615
KEYWORDS
SOURCE Shuttle vector pRHB17
ORGANISM Shuttle vector pRHB17
artificial sequences; vectors.

REFERENCE 1
AUTHORS Birmucci, M.G., Cheng, Q., Kostichka, K.N. and Tomb, J.F.
TITLE Rhodococcus cloning and expression vectors
JOURNAL Patent: WO 02055709-A 6 18-JUL-2002;
E.I. DU PONT DE NEMOURS AND COMPANY (US)

FEATURES
Source
1..11241
/organism="Shuttle vector pRHB17"
/mol_type="genomic DNA"
/db_xref="taxon:214938"

BASE COUNT 2517 a 3109 c 2945 g 2670 t
ORIGIN

Query Match 100.0%; Score 1140; DB 6; Length 11241;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1140; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGACGACGATGATGCTGAAACCTTTCCGCAAGACCGGCTCCGTCCTGCTGTCG 60
Db 6641 ATGACGACGATGATGCTGAAACCTTTCCGCAAGACCGGCTCCGTCCTGCTGTCG 6582
QY 61 TTCCGATAGCCGCGGATCCGCGACGAACTGGGACCCCAACTTCAACATCCACGCTCA 120
Db 6581 TTCCGATAGCCGCGGATCCGCGACGAACTGGGACCCCAACTTCAACATCCACGCTCA 6522
QY 121 GAAACATTTAAGCGCTGTGCGCGCGCGCAATTTCTGCGTGAAGCGGTGACCATTTGTCAAC 180
Db 6521 GAAACATTTAAGCGCTGTGCGCGCGCGCAATTTCTGCGTGAAGCGGTGACCATTTGTCAAC 6462
QY 181 GGTCCGAAAGGTTCTGATGATGAGAGGCTTCTGCTGCGGAAAGGCTGATCTGCCCC 240
Db 6461 GGTCCGAAAGGTTCTGATGATGAGAGGCTTCTGCTGCGGAAAGGCTGATCTGCCCC 6402
QY 241 TGCTGTGGGGGAAAGTGGTGCACATGTCGACGAAATTTCTCAAGTTGTCAT 300
Db 6401 TGCTGTGGGGGAAAGTGGTGCACATGTCGACGAAATTTCTCAAGTTGTCAT 6342
QY 301 CAACCTCGGAGCTGATCTGTGCGATGTCGATGACATGACATGCGCATACAGCTGTGAG 360
Db 6341 CAACCTCGGAGCTGATCTGTGCGATGTCGATGACATGACATGCGCATACAGCTGTGAG 6282
QY 361 CGGCTTCAAGCATATGATGATGAGCTTTTGGAGGCTTGGAAAGGCTGCAACAGGCTGT 420
Db 6281 CGGCTTCAAGCATATGATGATGAGCTTTTGGAGGCTTGGAAAGGCTGCAACAGGCTGT 6222
QY 421 CGTTGGCGTACGGAAGCGTGAATGTAAGCGTGGCGGATGACGATGACGCGCTGTAATTC 480
Db 6221 CGTTGGCGTACGGAAGCGTGAATGTAAGCGTGGCGGATGACGATGACGCGCTGTAATTC 6162
QY 481 ACTCAAGGAAAAAGCGCTGACGCTGCACTTCAAGCGCTACATGTTCAAGTGTGAC 540
Db 6161 ACTCAAGGAAAAAGCGCTGACGCTGCACTTCAAGCGCTACATGTTCAAGTGTGAC 6102

QY 541 GTGAGTGAAGCAATCTCGAATCTCTCGGATCCGATGTTTCGATCGGATGACCTTCCAAA 600
Db 6101 GTGAGTGAAGCAATCTCGAATCTCTCGGATCCGATGTTTCGATCGGATGACCTTCCAAA 6042
QY 601 CTGCTATCTCTGGATTTGGCTGCGCAGTACGTAATTCGGGTGCTCGATGTACGAAAG 660
Db 6041 CTGCTATCTCTGGATTTGGCTGCGCAGTACGTAATTCGGGTGCTCGATGTACGAAAG 5982
QY 661 ATCGGCGTGAAGCTGATCAAGTTCTGCTGCGATCTGAAGAAATTCGATCTCGGCTT 720
Db 5981 ATCGGCGTGAAGCTGATCAAGTTCTGCTGCGATCTGAAGAAATTCGATCTCGGCTT 5922
QY 721 GGTATGAGGTTGTGTATGATGAGGAGGATCCAGAGGTTGAACTGTGCGCGAATTT 840
Db 5921 GGTATGAGGTTGTGTATGATGAGGAGGATCCAGAGGTTGAACTGTGCGCGAATTT 5862
QY 841 GAAATCGCTGTTGATGACGATGAGGAGGATCCAGAGGTTGAACTGTGCGCGAATTT 840
Db 5861 GAAATCGCTGTTGATGACGATGAGGAGGATCCAGAGGTTGAACTGTGCGCGAATTT 5802
QY 841 GAGTTTGGTTCGATGAGGAGGATCCAGAGGTTGATGAGGAGGATTCGCGCGAGCT 900
Db 5801 GAGTTTGGTTCGATGAGGAGGATCCAGAGGTTGATGAGGAGGATTCGCGCGAGCT 5742
QY 901 GGTCTTGGGCGGAACTTAACGATGCTCAATCTGATGACGAGAAATCTGCGCGGCTC 960
Db 5741 GGTCTTGGGCGGAACTTAACGATGCTCAATCTGATGACGAGAAATCTGCGCGGCTC 5682
QY 961 ATGCTTGGGATCTTCCGCGCGAGTCTGATGATGATTCGAGCTTGGCGCCTTACGTC 1020
Db 5681 ATGCTTGGGATCTTCCGCGCGAGTCTGATGATGATTCGAGCTTGGCGCCTTACGTC 5622
QY 1021 TTGGCGGAGATCTCGGACCTGTCGAAAGCTGGCGCGAATTGGGAAAATCTTCGTATAC 1080
Db 5621 TTGGCGGAGATCTCGGACCTGTCGAAAGCTGGCGCGAATTGGGAAAATCTTCGTATAC 5562
QY 1081 TTGCATTATGATGATGCGGAGCGGAGTGGCGGCGCGCGATTAATATGCTTGGCAAGTGA 1140
Db 5561 TTGCATTATGATGATGCGGAGCGGAGTGGCGGCGCGCGATTAATATGCTTGGCAAGTGA 5502

RESULT 6
AY180162/c

LOCUS AY180162 11241 bp DNA circular SYN 01-MAY-2003
DEFINITION Shuttle vector pRHB17, complete sequence.
ACCESSION AY180162
VERSION AY180162.1 GI:30313707
KEYWORDS

SOURCE Shuttle vector pRHB17
ORGANISM Shuttle vector pRHB17
artificial sequences; vectors.

REFERENCE 1 (bases 1 to 11241)
AUTHORS Kostichka, K., Tao, L., Birmucci, M., Tomb, J.-F., Nagarajan, V. and Cheng, Q.*

TITLE A small cryptic plasmid from Rhodococcus erythropolis:
characterization and utility for gene expression
JOURNAL Unpublished
2 (bases 1 to 11241)

REFERENCE Kostichka, K., Tao, L., Birmucci, M., Tomb, J.-F., Nagarajan, V. and Cheng, Q.
AUTHORS

TITLE Direct Submission
JOURNAL Submitted (14-NOV-2002) CRØD, E. I. Dupont de Nemours Inc.,
Experimental Station, Wilmington, DE 19880-0328, USA

FEATURES
Location/Qualifiers
1..11241
/organism="Shuttle vector pRHB17"
/mol_type="genomic DNA"
/db_xref="taxon:214938"
/focus

/note="Rhodococcus - Escherichia coli shuttle vector; Ssp
I linearize pAN12 cloned into the PvuII site of pBR328"
join(9925..11241,1..3590)
/organism="Cloning vector pBR328"

Accession	Sequence	Position
Db	5861 GAAACCGCTGTGATGCACGTGGGCGGATCCACAAGCCTTGGAACTGTGGCAGATTT	580
QY	841 GAGTTTGATTGCATGGGACGTGGGCAATCGGGTGTCCCGTGATTTGGTGCCGAGCT	900

REFERENCE
1 (pages 1 to 20289)
Muzny, D., Marie, Metzker, M., Lee, Abramson, S., Adams, C., Alder, J.,
Muzny, D., Marie, Metzker, M., Lee, Abramson, S., Adams, C., Alder, J.,

Ayala-Bechei, V., Aoyagi, A., Ayodele, M., Bacc, E., Baden, H.,
Baldwin, D., Bandaranaike, D., Barber, M., Barneshead, M., Bernhmed, F.,
Bismail, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,
Bryan, N., Buhey, C., Burck, P., Burrell, K., Calderon, E.,
Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Cencer, A.,
Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J.,
Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L.,
Davis, M. L., Denis, C., Davy-carroll, L., De Ande, C., Dederich, D.,
Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K.,
Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Evans, K.,
Egan, A., Escotto, M., Eugene, C., Evans, C. A., Falls, T., Fan, G.,
Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,
Fraser, C. M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M.,
Gebreorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guayara, M.,
Gunaratne, P., Haaland, W., Hamli, C., Hamilton, C., Hamilton, K.,
Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J.,
Hernandez, R., Hines, S., Hladun, S. L., Hodgson, A., Hogues, M.,
Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A.,
Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolyvet, A.,
Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C.,
Kowis, C., Kraft, C. L., Lebow, H., Levay, J., Lewis, L., Li, Z., Liu, J.,
Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J.,
Lorenzshewa, L., Loulseged, H., Lozado, R. J., Lu, X., Ma, J.,
Maheshwari, M., Mahindaratne, M., Mahmood, M., Malloy, K., Mangum, A.,
Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E.,
Mawhinney, S., McLeod, M. P., McNeill, T. Z., Meenen, E.,
Milosavljevic, A., Miner, G., Munja, B., Montemayor, J., Moore, S.,
Morgan, M., Morris, K., Morris, S., Mundasa, M., Murphy, M., Nair, L.,
Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S.,
Nwoakemah, O., Okunwo, G., Olampunagoon, A., Pal, S., Parks, K.,
Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C.,
Plopper, F., Poindexter, A., Popovic, D., Prims, E., Pu, L., L.,
Puzo, M., Quiroz, J., Rachlin, B., Reeves, K., Reider, M. A., Reich, R.,
Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F.,
Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S. U.,
Sanders, W., Savery, G., Scherer, S., Scott, G., Shatsman, S., Shen, H.,
Shetty, J., Shwartsbeyn, A., Slisson, I., Sitter, C. D., Smjs, D.,
Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, C.,
Steimle, M., Strong, R., Sutton, A., Svaltek, A., Tabori, P., Taylor, C.,

Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K., Vales, R., Vera, V., Villasana, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, P., Williams, G., Wilson, R., Wleczek, R., Woodson, H., Wotley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, S., Zhao, S., Dunn, D., von Niederhausern, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O., Weinstock, G. and Gibbs, R.A.

Direct Submission
Unpublished
2 (bases 1 to 202897)
Rat Genome Sequencing Consortium.

REFERENCE
TITLE
JOURNAL
AUTHORS
SUBMITTED (12-SEP-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 202897)
Rat Genome Sequencing Consortium.

COMMENT
Direct Submission
Submitted (15-NOV-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
On Nov 15, 2002 this sequence version replaced gi:22795096.
The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: KBRK
Center clone name: CH230-281C3
----- Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 176284 bases at least Q40
Consensus quality: 178546 bases at least Q30
Consensus quality: 180221 bases at least Q20
Estimated insert size: 183234; sum-of-contigs estimation
Quality coverage: 7x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 1 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
1 202897: contig of 202897 bp in length.
Location/Qualifiers
1..202897
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-281C3"
2..752
/note="clone boundary
clone_end:Sp6
site:"

misc_feature
misc_feature

end_sequence:B256844"
87041..88414
/note="wgs contig"
complement(198268..199155)
/note="clone boundary
clone_end:17
site:
end_sequence:B256842"
200795..202897
/note="wgs_end-extension
clone_end:17"
misc_feature
BASE COUNT 45958 a 41690 c 43148 g 51167 t 20934 others
ORIGIN
Query Match 1.9%; Score 22; DB 2; Length 202897;
Best Local Similarity 100.0%; Pred. No. 1.9;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 765 CAACGTCGACCCCTGGAAATC 786
Db 124096 CAACGTCGACCCCTGGAAATC 124075
AC123219 224046 bp DNA linear HTG 22-SEP-2002
Rattus norvegicus clone CH230-175M19, *** SEQUENCING IN PROGRESS
*** 6 unordered pieces.
AC123219
AC123219.3 GI:23195026
HTG: HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE
Rattus norvegicus (Norway rat)
ORGANISM
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 224046)
REFERENCE
AUTHORS
Muzny, D., Marie, Metzger, M., Lee, Adamson, S., Adams, C., Alder, J., Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D., Anylebech, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H., Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F., Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M., Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E., Cardenas, V., Carter, K., Cavazos, I., Caesar, H., Center, A., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J., Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L., Davila, M.L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D., Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K., Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Baves, K., Egan, A., Escotto, M., Eugene, C., Evans, C.A., Falls, T., Fan, G., Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P., Fraser, C.M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M., Gebregeorgis, E., Geer, K., Gill, R., Grady, M., Guerra, M., Guevara, W., Gunaratne, P., Haaland, M., Hamill, C., Hamilton, C., Hamilton, K., Harvey, Y., Haylak, P., Hawes, J., Henderson, N., Hernandez, J., Hernandez, R., Hines, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A., Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A., Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A., Karpathy, S., Kelly, S., Kelly, S., Khan, S., King, R., Kovar, C., Kowis, C., Kraft, C.L., Lebow, H., Lewan, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Lorenshewal, L., Louisedge, H., Lozada, R.J., Lu, X., Ma, J., Maheshwari, M., Mahindartine, M., Mahmoud, M., Malloy, K., Mangum, A., Mangum, B., Mapa, P., Martin, K., Martin, R., Martinez, B., Mawhney, S., McLeod, M.P., McNeill, T.Z., Meenan, E., Milosavljevic, A., Miner, G., Minga, B., Montemayor, J., Moore, S., Morgan, N., Morris, K., Morris, S., Munidasa, M., Murphy, M., Naik, L., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C., Plopper, F., Poindexter, A., Popovic, D., Primus, B., Pu, L., L., Puzo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F.,

REFERENCE	AUTHORS	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM	
1	Mazny,D,Marle, Metzker,M, Lee, Abramzon,S, Adams,C, Alder,J, Allen,C, Allen,H, Aishrooks,S, Antin,A, Angiano,D, Anyalebechi,V, Aoyagi,A, Ayodeji,M, Baca,E, Baden,H, Baldwin,D, Bandaranaike,D, Barber,M, Barnstead,M, Benahmed,F, Biswalto,K, Blair,J, Blankenburg,K, Blyth,P, Brown,M, Bryant,N, Buhay,C, Burch,P, Butrell,K, Calderon,E, Cadenas,V, Carter,K, Cavazos,I, Ceasar,H, Chen,A, Chacko,K, Chavez,D, Chen,G, Chen,R, Chen,Y, Chen,Z, Chu,J, Cleveland,C, Cockrell,R, Cox,C, Coyle,M, Cree,A, D'Souza,L, David,M,L, Davis,C, Day-carroll,L, De Anda,C, Dederich,D, Delgado,O, Denson,S, Deramo,C, Ding,Y, Dim,H, Divya,K, Draper,H, Dugan-Rocha,S, Dunn,A, Durbin,K, Duval,B, Eaves,K, Egan,A, Escotto,M, Eugene,C, Evans,C,A, Falls,T, Fan,G, Fernandez,S, Finley,M, Flagg,N, Forbes,L, Foster,M, Foster,P, Fraser,C,M, Gabisi,A, Ganta,R, Garcia,A, Garner,T, Garza,M, Gebregeorgis,E, Geer,K, Gill,R, Grady,M, Guerra,W, Guevara,W, Guvaratne,P, Haaland,M, Hamil,C, Hamilton,C, Hamilton,K, Haney,Y, Havlik,P, Hawes,A, Henderson,N, Hernandez,J, Hernandez,R, Hines,S, Hladun,S,L, Hodgson,A, Hogues,M, Hollins,B, Howells,S, Huliy,K,S, Hume,J, Idlebird,D, Jackson,A, Jackson,L, Jacob,L, Jiang,H, Johnson,Z, Johnson,R, Jolyvet,A, Karpathy,S, Kelly,S, Kelly,S, Khan,Z, King,L, Kovar,C, Kowis,C, Kraft,C,L, Lebow,H, Levan,U, Lewis,L, Li,Z, Liu,J, Liu,J, Liu,W, Liu,Y, London,P, Longacre,S, Lopez,J, Lorenshewa,L, Louisedge,H, Lozado,R,J, Lu,X, Ma,J, Mahalingam,M, Mahindartine,M, Mahmoud,M, Malloy,K, Mangum,A, Margues,B, Mapua,P, Martin,K, Martin,R, Martinez,E,	AC103497	AC103497	AC103497	AC103497	AC103497	AC103497
2	228864 bp	DNA	linear	HTG 10-MAY-2002			
3	100.0%	Pred. No. 1.9%					
4	0	Mismatches 0	Indels 0	Gaps 0			
5	765	CAACCGTGCACCCCTGGGAAATC	786				
6	196332	CAACCGTGCACCCCTGGGAAATC	196353				
7	Query Match	1.9%	Score 22	DB 2	Length 224046		
8	Best Local Similarity	100.0%	Pred. No. 1.9%				
9	Matches 22	Conservative 0	Mismatches 0	Indels 0	Gaps 0		
10	misc_feature	/note="wgs_end_extension clone_end:T7"					
11	misc_feature	/note="wgs_end_extension clone_end:T7"					
12	misc_feature	/note="wgs_end_extension clone_end:T7"					
13	misc_feature	/note="wgs_end_extension clone_end:T7"					
14	misc_feature	/note="wgs_end_extension clone_end:T7"					
15	misc_feature	/note="wgs_end_extension clone_end:T7"					
16	misc_feature	/note="wgs_end_extension clone_end:T7"					
17	misc_feature	/note="wgs_end_extension clone_end:T7"					
18	misc_feature	/note="wgs_end_extension clone_end:T7"					
19	misc_feature	/note="wgs_end_extension clone_end:T7"					
20	misc_feature	/note="wgs_end_extension clone_end:T7"					
21	misc_feature	/note="wgs_end_extension clone_end:T7"					
22	misc_feature	/note="wgs_end_extension clone_end:T7"					
23	misc_feature	/note="wgs_end_extension clone_end:T7"					
24	misc_feature	/note="wgs_end_extension clone_end:T7"					
25	misc_feature	/note="wgs_end_extension clone_end:T7"					
26	misc_feature	/note="wgs_end_extension clone_end:T7"					
27	misc_feature	/note="wgs_end_extension clone_end:T7"					
28	misc_feature	/note="wgs_end_extension clone_end:T7"					
29	misc_feature	/note="wgs_end_extension clone_end:T7"					
30	misc_feature	/note="wgs_end_extension clone_end:T7"					
31	misc_feature	/note="wgs_end_extension clone_end:T7"					
32	misc_feature	/note="wgs_end_extension clone_end:T7"					
33	misc_feature	/note="wgs_end_extension clone_end:T7"					
34	misc_feature	/note="wgs_end_extension clone_end:T7"					
35	misc_feature	/note="wgs_end_extension clone_end:T7"					
36	misc_feature	/note="wgs_end_extension clone_end:T7"					
37	misc_feature	/note="wgs_end_extension clone_end:T7"					
38	misc_feature	/note="wgs_end_extension clone_end:T7"					
39	misc_feature	/note="wgs_end_extension clone_end:T7"					
40	misc_feature	/note="wgs_end_extension clone_end:T7"					
41	misc_feature	/note="wgs_end_extension clone_end:T7"					
42	misc_feature	/note="wgs_end_extension clone_end:T7"					
43	misc_feature	/note="wgs_end_extension clone_end:T7"					
44	misc_feature	/note="wgs_end_extension clone_end:T7"					
45	misc_feature	/note="wgs_end_extension clone_end:T7"					
46	misc_feature	/note="wgs_end_extension clone_end:T7"					
47	misc_feature	/note="wgs_end_extension clone_end:T7"					
48	misc_feature	/note="wgs_end_extension clone_end:T7"					
49	misc_feature	/note="wgs_end_extension clone_end:T7"					
50	misc_feature	/note="wgs_end_extension clone_end:T7"					
51	misc_feature	/note="wgs_end_extension clone_end:T7"					
52	misc_feature	/note="wgs_end_extension clone_end:T7"					
53	misc_feature	/note="wgs_end_extension clone_end:T7"					
54	misc_feature	/note="wgs_end_extension clone_end:T7"					
55	misc_feature	/note="wgs_end_extension clone_end:T7"					
56</							

to be correct as given, however the sizes of the gaps between them are based on estimates that have been provided by the submitter. This sequence will be replaced by the finished sequence as soon as it is available and the accession number will be preserved.

* NOTE: This is a 'working draft' sequence.

* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.

FEATURES

Location/Qualifiers

1. 141405

/organism="Oryza sativa (japonica cultivar-group)"

/mol_type="genomic DNA"

/cultivar="Nipponbare"

/db_xref="taxon:39947"

/chromosome="8"

/clone="OSJNBa0036R18"

BASE COUNT 41811 a 28829 c 29018 g 41609 t 138 others

ORIGIN

Query Match 1.8%; Score 21; DB 2; Length 141405;

Best Local Similarity 100.0%; Pred. No. 7.4;

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1054 GCGACTTGCGAAATCTTCGT 1074

Db 22729 GCGACTTGCGAAATCTTCGT 22749

RESULT 11

AP005799

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

JOURNAL

JOURNAL

JOURNAL

JOURNAL

JOURNAL

JOURNAL

JOURNAL

JOURNAL

JOURNAL

JOURNAL

JOURNAL

JOURNAL

JOURNAL

JOURNAL

JOURNAL

JOURNAL

JOURNAL

JOURNAL

AP005799 155547 bp DNA linear HTG 03-OCT-2002
Oryza sativa (japonica cultivar-group) chromosome 8 clone
OSJNBa0012K14, *** SEQUENCING IN PROGRESS ***.

AP005799.1 GI:23496503
HTG; HTGS_PHA52.
Oryza sativa (japonica cultivar-group)
Oryza sativa (japonica cultivar-group)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliopsida; Liliopsida; Poales; Poaceae;
Erihartoideae; Oryzaceae; Oryza.

1
Sasaki, T., Matsumoto, T. and Katayose, Y.
Oryza sativa nipponbare (GA3) genomic DNA, chromosome 8, BAC
clone:OSJNBa0012K14
Published Only in Database (2002)
2 (bases 1 to 155547)
Sasaki, T., Matsumoto, T. and Katayose, Y.
Direct Submission
Submitted (02-OCT-2002) Takuji Sasaki, National Institute of
Agrobiological Sciences, Rice Genome Research Program, Kannondai
2-1-2, Tsukuba, Ibaraki 305-8602, Japan
E-mail:tsasaki@nias.affrc.go.jp, URL:http://rgrp.dna.affrc.go.jp/
Tel:81-298-7441, Fax:81-298-38-7468)
NOTE: It currently consists of 1 contigs. Gaps between the contigs
are represented as runs of N. The order of the pieces is believed
to be correct as given, however the sizes of the gaps between them
sequence will be replaced by the finished sequence as soon as it is
available and the accession number will be preserved.
* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.

FEATURES

Location/Qualifiers

1. 155547

/organism="Oryza sativa (japonica cultivar-group)"

/mol_type="genomic DNA"

/cultivar="Nipponbare"

/db_xref="taxon:39947"

/chromosome="8"

/clone="OSJNBa0012K14"

BASE COUNT 45247 a 32175 c 32490 g 45635 t

ORIGIN

Query Match 1.8%; Score 21; DB 2; Length 155547;

Best Local Similarity 100.0%; Pred. No. 7.3;

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1054 GCGACTTGCGAAATCTTCGT 1074

Db 147013 GCGACTTGCGAAATCTTCGT 147033

RESULT 12

AX548660

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

JOURNAL

JOURNAL

JOURNAL

JOURNAL

JOURNAL

JOURNAL

JOURNAL

JOURNAL

JOURNAL

JOURNAL

JOURNAL

JOURNAL

JOURNAL

JOURNAL

JOURNAL

JOURNAL

JOURNAL

JOURNAL

JOURNAL

JOURNAL

JOURNAL

JOURNAL

JOURNAL

JOURNAL

JOURNAL

JOURNAL

JOURNAL

JOURNAL

AX548660 20 bp DNA linear PAT 27-NOV-2002
Sequence 19 from Patent WO02055709.
AX548660
AX548660.1 GI:25813628
synthetic construct
synthetic construct
artificial sequences.

1
Bramucci, M.G., Cheng, Q., Kostichka, K.N. and Tomb, J.F.
Rhodococcus cloning and expression vectors
Patent: WO 02055709-A 19 18-JUL-2002;
E.I. DU PONT DE NEMOURS AND COMPANY (US)
Location/Qualifiers
1. 20
/organism="synthetic construct"
/mol_type="genomic DNA"
/db_xref="taxon:32630"
/note="Primer"

BASE COUNT 6 a 5 c 3 g 6 t

ORIGIN

Query Match 1.8%; Score 20; DB 6; Length 20;

Best Local Similarity 100.0%; Pred. No. 39;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1119 GATTAATATCGTTTCGCACT 1138

Db 20 GATTAATATCGTTTCGCACT 1

RESULT 13

AX049359

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

JOURNAL

JOURNAL

JOURNAL

JOURNAL

JOURNAL

JOURNAL

JOURNAL

JOURNAL

AX049359 167 bp DNA linear PAT 12-JAN-2001
Sequence 29 from Patent WO0071709.
AX049359
AX049359.1 GI:12226115
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

1
Gires, B., Gasnier, B., Sagne, C., el Mestikawy, S. and Hamon, M.
Polypeptides, vesicular carriers of glutamate and gaba
Patent: WO 0071709-A 29 30-NOV-2000;
INSTITUT NATIONAL DE LA SANTE ET DE LA RECHERCHE MEDICALE; (INSERM)
(FR)

FEATURES

Location/Qualifiers

1. 167

/organism="Homo sapiens"

/mol_type="genomic DNA"

/db_xref="taxon:9606"

BASE COUNT 40 a 51 c 37 g 39 t

ORIGIN

Query Match 1.8%; Score 20; DB 6; Length 167;
Best Local Similarity 100.0%; Pred. No. 36;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 324 GATGGTACGATGACCATGC 343
Db 81 GATGGTACGATGACCATGC 62

RESULT 14
AY227111/c 1402 bp mRNA linear PRI 12-FEB-2003

LOCUS DEFINITION Homo sapiens clone IMAGE:2549054 truncated transmembrane transport protein (SLC36A1) mRNA, complete cds; alternatively spliced.

AY227111
ACCESSION protein (SLC36A1) mRNA, complete cds; alternatively spliced.
VERSION AY227111.1 GI:28372397
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiindae; Homo.

REFERENCE 1 (bases 1 to 1402)
AUTHORS Bermingham, J.R. Jr. and Pennington, J.
TITLE Genome organization and expression of the SLC36a cluster of amino acid transport protein genes
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1402)
AUTHORS Bermingham, J.R. Jr. and Pennington, J.
TITLE Direct Submission
JOURNAL Submitted (28-JAN-2003) McLaughlin Research Institute, 1520 23rd Street South, Great Falls, MT 59405, USA

FEATURES
source 1.1420
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/chromosome="5"
/map="5q33"
/clone="IMAGE:2549054"
1.1402
/gene="SLC36A1"
/note="LYAAT-1"
92..95
variation /gene="SLC36A1"
/note="apparently deleted and replaced by a 62 nt sequence in the genomic sequence"
/replace=""
144..1304
CDS /gene="SLC36A1"
/note="alternatively spliced; uses alternative 3' exon that results in C-terminal truncation; transmembrane protein"
/codon_start=1
/product="truncated transmembrane transport protein"
/protein_id="AAO37091.1"
/db_xref="GI:28372398"
/translation="MSTQRLNEDYHDYSDVSPESPSEGLNNTLSPGSYQRFQOS NSTWPTQTLHLKNGITGILGIPLVNAGIVMGIPISLIITIGTAVAGMGIVYCA HHRCPELINSFVYDGTWYVGLSSCSLRHRAHGRVYVFLVYDGLGCVYFA FLADNFKQVTEAANGTNNCHNNEVILPTMDSRLYMLSPFLVLFIRNLRLS IFSLLNTMLVSLVMIYQFIVQRPDPSPHLPLVAVKTYPLPFFGTAISFFGIQVLI PLNNKMDPRKRPPLIYLIGMIVTILYISGLGIVQFGANIQGSITLNDPMCMILYOS VKLIYSIGIFFTYALQFVYPAEIIIPFFVSRAPEHCELVVDLFEVRLVCLT"

BASE COUNT 312 a 393 c 352 g 345 t

ORIGIN

Query Match 1.8%; Score 20; DB 9; Length 1402;
Best Local Similarity 100.0%; Pred. No. 33;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 324 GATGGTACGATGACCATGC 343
Db 1046 GATGGTACGATGACCATGC 1027

RESULT 15
AY227112/c 1420 bp mRNA linear PRI 12-FEB-2003

LOCUS

DEFINITION Homo sapiens clone DKKFZp434G1123 transmembrane transport protein (SLC36A1) mRNA, partial cds.
AY227112
ACCESSION AY227112
VERSION AY227112.1 GI:28372399
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiindae; Homo.

REFERENCE 1 (bases 1 to 1420)
AUTHORS Bermingham, J.R. Jr. and Pennington, J.
TITLE Genome organization and expression of the SLC36a cluster of amino acid transport protein genes
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1420)
AUTHORS Bermingham, J.R. Jr. and Pennington, J.
TITLE Direct Submission
JOURNAL Submitted (28-JAN-2003) McLaughlin Research Institute, 1520 23rd Street South, Great Falls, MT 59405, USA

FEATURES
source 1.1420
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/chromosome="5"
/map="5q33"
/clone="DKKFZp434G1123"
/note="derived from dbEST AL043182"
1.1420
/gene="SLC36A1"
/note="LYAAT-1"
1..942
CDS /gene="SLC36A1"
/note="transmembrane transport protein"
/product="transmembrane transport protein"
/protein_id="AAO37092.1"
/db_xref="GI:28372400"
/translation="DNFKQVTEAANGTNNCHNNEVILPTMDSRLYMLSPFLVLI LVEIRNLRLSIFSLANTMLVSLVMIYQFIVQRPDPSPHLPLVAVKTYPLPFFGTA IFSFEGIGMVPLNNKMDPRKRPPLIYLIGMIVTILYISGLGIVQFGANIQGSIT LNPFCMILYOSVKLIYSIGIFFTYALQFVYPAEIIIPFFVSRAPEHCELVVDLFEVRLV LVCILCITLAILPRLDVLSIGVSSSLALIIPLPLEVTTFYSRGMSPLTIFKDAL ISILGFVGVVGTGYBALYELLIOPSNAPIFINSTCAFI"

BASE COUNT 281 a 456 c 315 g 368 t

ORIGIN

Query Match 1.8%; Score 20; DB 9; Length 1420;
Best Local Similarity 100.0%; Pred. No. 33;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 324 GATGGTACGATGACCATGC 343
Db 414 GATGGTACGATGACCATGC 395

Search completed: November 8, 2003, 22:29:30
Job time : 2941 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Comugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 8, 2003, 20:05:57 ; Search time 281 Seconds
(without alignments)
10951.457 Million cell updates/sec

Title: US-10-007-527A-1

Perfect score: 1140
Sequence: 1 atgaccagcgtcaagtcgtga.....taatatcggtcgaagtga 1140

Scoring table: OLIGO NUC
Gapop 60.0 , Gapext 60.0

Searched: 2552756 segs, 1349719017 residues

Word size : 0

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-Processing: Listing first 45 summaries

Database : N_Geneseq_19Jun03.*

1: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:*
2: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:*
3: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:*
4: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:*
5: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:*
6: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:*
7: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:*
8: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:*
9: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:*
10: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:*
11: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:*
12: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:*
13: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:*
14: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:*
15: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:*
16: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:*
17: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:*
18: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:*
19: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:*
20: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:*
21: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:*
22: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:*
23: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:*
24: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:*
25: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysts of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1140	100.0	1140	24	ABQ76122
2	1140	100.0	6334	24	ABQ76124
3	1140	100.0	9652	24	ABQ76126
4	1140	100.0	11241	24	ABQ76125
5	20	1.8	20	24	ABQ76138
6	20	1.8	157	23	ABL57927
7	20	1.8	739	23	AAS90653
8	20	1.8	947	23	ABV23244

C 9	20	1.8	947	23	ABV29088	Human prostate exp
C 10	20	1.8	1428	23	ABL57930	Human VG51 coding
C 11	20	1.8	1501	25	ABX70643	Human cDNA encoding
C 12	20	1.8	1513	25	ABX70642	Human cDNA encoding
C 13	20	1.8	1597	25	ABX70783	Novel human cDNA s
C 14	20	1.8	1771	23	AAS90655	DNA encoding novel
C 15	20	1.8	2057	25	ABZ80236	Human encoding novel
C 16	20	1.8	2807	25	ABX12019	Human titandorin 3
C 17	20	1.8	3114	24	AAI44674	Human transporter and 1
C 18	20	1.8	249999	25	ABZ80229	Human transmembrane
C 19	19	1.7	332	21	AAC72811	Single nucleotide
C 20	19	1.7	476	25	ABX22328	Human GDP-mannose
C 21	19	1.7	1209	24	ABV78077	Hypoxia-regulated
C 22	19	1.7	1333	22	AAH34705	Human colon cancer
C 23	19	1.7	3810	24	ABN97214	Gene #3712 used to
C 24	19	1.7	4515	23	ABL07083	Drosophila melanog
C 25	19	1.7	4578	10	AAN92192	cDNA sequence of c
C 26	19	1.7	4579	11	AAO62827	L-A virus genome.
C 27	19	1.7	4580	19	AAV05284	Saccharomyces cere
C 28	19	1.7	26370	23	ABL07082	Rhodococcus AN12 d
C 29	18	1.6	20	24	ABQ76139	Rhodococcus AN12 d
C 30	18	1.6	282	25	ABX25710	Human GDP-mannose
C 31	18	1.6	399	22	AAS35882	Human cardiovascular
C 32	18	1.6	440	22	AAS35179	Human cardiovascular
C 33	18	1.6	608	21	AAFO7379	Fusarium venenatum
C 34	18	1.6	619	23	AAV73050	DNA encoding novel
C 35	18	1.6	807	17	AAV15535	DNA encoding rat c
C 36	18	1.6	982	19	AAV52040	Helicobacter polyP
C 37	18	1.6	1065	23	AAV71237	DNA encoding novel
C 38	18	1.6	1095	23	AAS52556	E. coli DNA for ce
C 39	18	1.6	1659	23	AAS82835	DNA encoding novel
C 40	18	1.6	4104	24	ABZ35783	Human IGF1R polyu
C 41	18	1.6	4104	24	ABX10026	Human IGF1R DNA fr
C 42	18	1.6	4104	24	ABV78207	Human IGF1R DNA SE
C 43	18	1.6	4104	24	ABL91748	Human polynucleoti
C 44	18	1.6	4383	25	ABZ71068	Mycobacterium tube
C 45	18	1.6	4575	15	AAQ73445	IGF-I receptor. H

ALIGNMENTS

RESULT 1
ABQ76122
ID ABQ76122 standard; DNA; 1140 BP.
XX
AC ABQ76122;
XX
DT 13-JUN-2003 (first entry)
XX
DE Rhodococcus AN12 replication protein Rep DNA.
XX
XX Plasmid stability protein; replication protein; ethylene forming enzyme;
KW carotenoid biosynthesis enzyme; isoprenoid; pyruvate decarboxylase;
KW polyhydroxyalkanoic acid synthase; PBA synthase; nitrile hydratase;
KW alcohol dehydrogenase; terpene synthase; cholesterol oxidase; gene;
KW shuttle vector; Rep; ds.
XX
OS Rhodococcus erythropolis.
XX
XX ~~MO200255709-A2.~~
XX
XX 18-JUL-2002.
XX
XX 12-DEC-2001; 2001MO-US47868.
XX
XX 12-DEC-2000; 2000US-254868P.
XX
XX (DUPO) DU PONT DE NEMOURS & CO E. I.
XX
XX ~~Branucci MG, Cheng Q, Kostichka KN, Tomb J,~~
XX ~~WFI, 2002-557827/59.~~

DR P-PSDB; ABB84278.

XX New nucleic acid molecule encoding replication protein/plasmid

PT stability protein, useful in cloning and expression vectors,

PT particularly shuttle vectors for expression of heterologous genes in

PT Rhodococcus species -

XX

PS Claim 2; Page 63-64; 96pp; English.

XX This invention describes a novel nucleic acid encoding a replication

CC protein or a plasmid stability protein. The product of the invention is

CC useful for expression of nucleic acid such as genes encoding enzymes

CC involved in the production of isoprenoid molecules, polyhydroxyalkanoic

CC acid (PHA) synthases, carotenoid biosynthesis enzymes, nitrile

CC hydratases, ethylene forming enzyme, pyruvate decarboxylase, alcohol

CC dehydrogenase, terpene synthases, and cholesterol oxidase in an

CC actinomycetals bacteria. The replication protein or plasmid stability

CC protein are useful in cloning and expression vectors and particularly in

CC shuttle vectors for the expression of homologous and heterologous genes

CC in Rhodococcus sp. This sequence encodes the Rhodococcus AN12 Rep

CC protein described in the disclosure of the invention.

XX

SQ Sequence 1140 BP; 241 A; 287 C; 347 G; 265 T; 0 other;

Query Match 100.0%; Score 1140; DB 24; Length 1140;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 1140; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGACGAGGCTAAGTCTGTAACACCTTTCCGGAAGACCGGCTCTCCCTCTGCTG 60

DB 1 ATGACGAGGCTAAGTCTGTAACACCTTTCCGGAAGACCGGCTCTCCCTCTGCTG 60

QY 61 TTCGATTAAGCGCGGCTATCCGCGACGACCTGCAACCTTCAAAATCAACCGTCA 120

DB 61 TTCGATTAAGCGCGGCTATCCGCGACGACCTGCAACCTTCAAAATCAACCGTCA 120

QY 121 GAAACATTTAAGCGCTGCGCGCGCGGATTTCTGCGTGAACGCTGACATTTGTCAAC 180

DB 121 GAAACATTTAAGCGCTGCGCGCGCGGATTTCTGCGTGAACGCTGACATTTGTCAAC 180

QY 181 GGTCCGAAAGGTTCTGATTCGGAAGGCTTCTCTGCGGAAAGGCTGATCTGCCCC 240

DB 181 GGTCCGAAAGGTTCTGATTCGGAAGGCTTCTCTGCGGAAAGGCTGATCTGCCCC 240

QY 241 TGTCTGCGGGAAGGTCGCTGCAATCGTGGCAGAGAAATTTCTCAATTTTGTGCTCAT 300

DB 241 TGTCTGCGGGAAGGTCGCTGCAATCGTGGCAGAGAAATTTCTCAATTTTGTGCTCAT 300

QY 301 CAATCGGGAAGGTCGATCTGTTGCGATGTCAGATGACATGCGCCATACAGCTGTGAC 360

DB 301 CAATCGGGAAGGTCGATCTGTTGCGATGTCAGATGACATGCGCCATACAGCTGTGAC 360

QY 361 CGGCTCCACGACCTATGATCTGTTGCGATGTCAGATGACATGCGCCATACAGCTGTGAC 420

DB 361 CGGCTCCACGACCTATGATCTGTTGCGATGTCAGATGACATGCGCCATACAGCTGTGAC 420

QY 421 CGTTGGGCTACGGAAGCTGAATGTAAGCTGCGACGGAATGCGGCGCTGTGAATC 480

DB 421 CGTTGGGCTACGGAAGCTGAATGTAAGCTGCGACGGAATGCGGCGCTGTGAATC 480

QY 481 ACTCAGGGAAGGTCGCTGCGACGTCACGCTTCAAGCTGATCTCAAGTGTGAC 540

DB 481 ACTCAGGGAAGGTCGCTGCGACGTCACGCTTCAAGCTGATCTCAAGTGTGAC 540

QY 541 GTGAGTGAAGACATCTCGAATCTTCTCGGATCGAATGTTGATCGGTGACTTCCAAA 600

DB 541 GTGAGTGAAGACATCTCGAATCTTCTCGGATCGAATGTTGATCGGTGACTTCCAAA 600

QY 601 CTGCTATCTCTGGGATTTGCTGCGGCTACATGTAATTCGGTGTCTCGATGTAAG 660

DB 601 CTGCTATCTCTGGGATTTGCTGCGGCTACATGTAATTCGGTGTCTCGATGTAAG 660

QY 661 ATGCGCGTGAAGTGAATCAAGTCTCGCTGCTGATCTGAAGAAATTTGATCTGGGCTT 720

DB 661 ATGCGCGTGAAGTGAATCAAGTCTCGCTGCTGATCTGAAGAAATTTGATCTGGGCTT 720

QY 721 GGTATGAGGTTGGTATGCGCGACGGAAGGTCGATGCAACCGTGACCCCTGG 780

DB 721 GGTATGAGGTTGGTATGCGCGACGGAAGGTCGATGCAACCGTGACCCCTGG 780

QY 781 GAAATCGCTGTTGATGACGATGCGGCGGATTCACAGCGTTGAACTGTGCGGAAATTT 840

DB 781 GAAATCGCTGTTGATGACGATGCGGCGGATTCACAGCGTTGAACTGTGCGGAAATTT 840

QY 841 GAGTTTGGTTGATGAGGACGTCGCGCAATCGCGTGTCTCGATGTCGCGGAACT 900

DB 841 GAGTTTGGTTGATGAGGACGTCGCGCAATCGCGTGTCTCGATGTCGCGGAACT 900

QY 901 GGTCTGGGCGAGAACTAAGATGCTCAAGTCTGTTGAGCAGAAATCTGCCCGGTC 960

DB 901 GGTCTGGGCGAGAACTAAGATGCTCAAGTCTGTTGAGCAGAAATCTGCCCGGTC 960

QY 961 ATGCTTGGGCGAGAACTAAGATGCTCAAGTCTGTTGAGCAGAAATCTGCCCGGTC 1020

DB 961 ATGCTTGGGCGAGAACTAAGATGCTCAAGTCTGTTGAGCAGAAATCTGCCCGGTC 1020

QY 1021 TTGCGCGAGATCTCTGCACTGTCGAAAGTGGCGGACTTGGGAAATCTTGTGATCAC 1080

DB 1021 TTGCGCGAGATCTCTGCACTGTCGAAAGTGGCGGACTTGGGAAATCTTGTGATCAC 1080

QY 1081 TTGCAATTCATGATTCGCGCGAGATGTCGCGCGCGGATTCGCGATTCGCGATTCG 1140

DB 1081 TTGCAATTCATGATTCGCGCGAGATGTCGCGCGCGGATTCGCGATTCGCGATTCG 1140

RESULT 2

ABQ76124/c

ID ABQ76124 standard; DNA; 6334 BP.

XX ABQ76124;

AC 13-JAN-2003 (first entry)

XX

DE Rhodococcus AN12 derived plasmid pAN12 DNA.

XX

KM Plasmid stability protein; replication protein; ethylene forming enzyme;

KM carotenoid biosynthesis enzyme; isoprenoid; pyruvate decarboxylase;

KM polyhydroxyalkanoic acid synthase; PHA synthase; nitrile hydratase;

KM alcohol dehydrogenase; terpene synthase; cholesterol oxidase;

KM shuttle vector; circular; ds.

XX

OS Rhodococcus erythropolis.

OS Synthetic.

XX

PN WO200255709-A2.

XX

PD 18-JUL-2002.

XX

PF 12-DEC-2001; 2001WO-US47868.

XX

PR 12-DEC-2000; 2000US-254868P.

XX

PA (DUPO) DU PONT DE NEMOURS & CO E. I.

XX

PI **BzAmuccl MG, Cheng Q, Kostichka KN, Tomb J;**

PI **BzAmuccl MG, Cheng Q, Kostichka KN, Tomb J;**

XX

DR WPI; 2002-557827/59.

XX

XX New nucleic acid molecule encoding replication protein/plasmid

PT stability protein, useful in cloning and expression vectors,

PT particularly shuttle vectors for expression of heterologous genes in

PT Rhodococcus species -

PS Claim 19; Page 68-71; 96pp; English.

XX This invention describes a novel nucleic acid encoding a replication

CC protein or a plasmid stability protein. The product of the invention is
 CC useful for expression of nucleic acid such as genes encoding enzymes
 CC involved in the production of isoprenoid molecules, polyhydroxyalkanoic
 CC acid (PHA) synthases, carotenoid biosynthesis enzymes, nitrile
 CC hydratases, ethylene forming enzyme, pyruvate decarboxylase, alcohol
 CC dehydrogenase, terpene synthases, and cholesterol oxidase in an
 CC actinomycetale bacteria. The replication protein or plasmid stability
 CC protein are useful in cloning and expression vectors and particularly in
 CC shuttle vectors for the expression of homologous and heterologous genes
 CC in *Rhodococcus* sp. This sequence represents the *Rhodococcus* AN12 derived
 CC plasmid pAN12 DNA described in the disclosure of the invention.

XX Sequence 6334 BP; 1386 A; 1807 C; 1681 G; 1460 T; 0 other:

Query Match 100.0%; Score 1140; DB 24; Length 6334;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 1140; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGACCAAGCTTAAGTGTGAACACCTTCCGGCAAGACCGGCTCCCGCTCTGCTGCG 60
 Db ATGACCAAGCTTAAGTGTGAACACCTTCCGGCAAGACCGGCTCCCGCTCTGCTGCG 2992
 QY 61 TTCGATTAAGCGCGGCAATCCGGGACGAACTGCGAACCCAACTTCAACCAATCACCACTCA 120
 Db TTCGATTAAGCGCGGCAATCCGGGACGAACTGCGAACCCAACTTCAACCAATCACCACTCA 2992
 QY 121 GAAACATTTTAACGCTTGGCCCGGCGAATTTCTGGCTGGAACGGTGTGACATTGTCAAC 180
 Db GAAACATTTTAACGCTTGGCCCGGCGAATTTCTGGCTGGAACGGTGTGACATTGTCAAC 2872
 QY 2931 GAAACATTTTAACGCTTGGCCCGGCGAATTTCTGGCTGGAACGGTGTGACATTGTCAAC 2872
 Db 181 GGTCCGAAAGCTTGTGATTCGAGAGCCCTTCTGCTCCGGAAGGCGCTGATTCGCCCC 240
 Db 2871 GGTCCGAAAGCTTGTGATTCGAGAGCCCTTCTGCTCCGGAAGGCGCTGATTCGCCCC 2812
 QY 241 TGCCTGTCGGGAAAGTCCGTCGCAATCGTCGCAACGAAATTTCTCAAGTTGTGTCAT 300
 Db TGCCTGTCGGGAAAGTCCGTCGCAATCGTCGCAACGAAATTTCTCAAGTTGTGTCAT 2752
 QY 2811 TGCCTGTCGGGAAAGTCCGTCGCAATCGTCGCAACGAAATTTCTCAAGTTGTGTCAT 2752
 Db 301 CAACTCGGGAATGATCTGTCGATGTCGATGTCGATGTCGATGTCGATGTCGATGTCGATG 360
 Db 2751 CAACTCGGGAATGATCTGTCGATGTCGATGTCGATGTCGATGTCGATGTCGATGTCGATG 2692
 QY 361 CGGCTCCACGACCTATGATGTCGATGTCGATGTCGATGTCGATGTCGATGTCGATGTCGATG 420
 Db CGGCTCCACGACCTATGATGTCGATGTCGATGTCGATGTCGATGTCGATGTCGATGTCGATG 2692
 QY 2691 CGGCTCCACGACCTATGATGTCGATGTCGATGTCGATGTCGATGTCGATGTCGATGTCGATG 2632
 Db 421 CGTGGCGTGAAGCAATGTCGATGTCGATGTCGATGTCGATGTCGATGTCGATGTCGATGTCGATG 480
 Db 2631 CGTGGCGTGAAGCAATGTCGATGTCGATGTCGATGTCGATGTCGATGTCGATGTCGATGTCGATG 2572
 QY 481 ACTACGGAAGAAACGCTGCGACGTCACGTCGTCGCTACTACTATGTCGATGTCGATGTCGATG 540
 Db ACTACGGAAGAAACGCTGCGACGTCACGTCGTCGCTACTACTATGTCGATGTCGATGTCGATG 2512
 QY 541 GTGAGTGAAGCAATCTCGAATCTCTTCTGAGATGTCGATGTCGATGTCGATGTCGATGTCGATG 600
 Db GTGAGTGAAGCAATCTCGAATCTCTTCTGAGATGTCGATGTCGATGTCGATGTCGATGTCGATG 2452
 QY 2511 GTGAGTGAAGCAATCTCGAATCTCTTCTGAGATGTCGATGTCGATGTCGATGTCGATGTCGATG 2452
 Db 601 CTGATATCTGCGAATTTGCTGCGCCTACTAGCTAATTTGGGGTCTGATGTCGATGTCGATGTCGATG 660
 Db 2451 CTGATATCTGCGAATTTGCTGCGCCTACTAGCTAATTTGGGGTCTGATGTCGATGTCGATGTCGATG 2392
 QY 720 ATCGGCGGTGAAGCTGATCAAGTTCTCGTGGCTATCTGACGAAATTTGCACTTGGCGTT 720
 Db ATCGGCGGTGAAGCTGATCAAGTTCTCGTGGCTATCTGACGAAATTTGCACTTGGCGTT 2332
 QY 2331 ATCGGCGGTGAAGCTGATCAAGTTCTCGTGGCTATCTGACGAAATTTGCACTTGGCGTT 2332
 Db 721 GGTATGAGAGTTGTTAGTGCACGCGAAGAAAGTGTGTCGATGTCGATGTCGATGTCGATGTCGATG 780
 Db GGTATGAGAGTTGTTAGTGCACGCGAAGAAAGTGTGTCGATGTCGATGTCGATGTCGATGTCGATG 2272
 QY 2272 GGTATGAGAGTTGTTAGTGCACGCGAAGAAAGTGTGTCGATGTCGATGTCGATGTCGATGTCGATG 2272
 Db 781 GAAATCGCTGTTGATGTCGATGTCGATGTCGATGTCGATGTCGATGTCGATGTCGATGTCGATG 840
 Db GAAATCGCTGTTGATGTCGATGTCGATGTCGATGTCGATGTCGATGTCGATGTCGATGTCGATG 2212
 Db 2271 GAAATCGCTGTTGATGTCGATGTCGATGTCGATGTCGATGTCGATGTCGATGTCGATGTCGATG 2212

QY 841 GAGTTTGGTTCAGTGGGACGTCGGGCAATCGCGTGGTCCCGTGAATTTGGCTCCCGAGCT 900
 Db 2211 GAGTTTGGTTCAGTGGGACGTCGGGCAATCGCGTGGTCCCGTGAATTTGGCTCCCGAGCT 2152
 QY 901 GGTCTTGGGCGCAAACTAACAGATGCTCAGATGTTGAGCAGAGAAATCTCCCGGCTC 960
 Db 2151 GGTCTTGGGCGCAAACTAACAGATGCTCAGATGTTGAGCAGAGAAATCTCCCGGCTC 2092
 QY 961 ATGCTTGGGCGCAATTTCCCGGCGGATGTCGATGTCGATGTCGATGTCGATGTCGATGTCGATG 1020
 Db 2091 ATGCTTGGGCGCAATTTCCCGGCGGATGTCGATGTCGATGTCGATGTCGATGTCGATGTCGATG 2032
 QY 1021 TTCGGGAGATTCCTGGAATCTGCGAAGCTGGGCGGCGACTTGGGAAATCTTGGTGAATCAC 1080
 Db 2031 TTCGGGAGATTCCTGGAATCTGCGAAGCTGGGCGGCGACTTGGGAAATCTTGGTGAATCAC 1972
 QY 1081 TTGCATTAATGATTTGCGCGCAGCGATGTCGCGCCCGCATATATCGTTTCGCAAGTGA 1140
 Db 1971 TTGCATTAATGATTTGCGCGCAGCGATGTCGCGCCCGCATATATCGTTTCGCAAGTGA 1912

RESULT 3

ABQ76126/c
 ID ABQ76126 standard; DNA; 9652 BP.

XX ABQ76126;

XX 13-JUN-2003 (first entry)

XX Plasmid pRHB171 DNA.

XX Plasmid stability protein; replication protein; ethylene forming enzyme;
 KW carotenoid biosynthesis enzyme; isoprenoid; pyruvate decarboxylase;
 KW polyhydroxyalkanoic acid synthase; PHA synthase; nitrile hydratase;
 KW alcohol dehydrogenase; terpene synthase; cholesterol oxidase;
 KW shuttle vector; circular; ds.

XX Synthetic.

XX WO200255709-A2.

XX 18-JUL-2002.

XX 12-DEC-2001; 2001WO-US47868.

XX 12-DEC-2000; 2000US-254868P.

XX (DUPO) DU PONT DE NEMOURS & CO E. I.

XX **Bramucci MG**, Cheng Q, Kostichka KN, Tomb J;

XX WPI; 2002-557827/59.

PT New nucleic acid molecule encoding replication protein/plasmid
 PT stability protein, useful in cloning and expression vectors,
 PT particularly shuttle vectors for expression of heterologous genes in
 PT *Rhodococcus* species

PS Claim 27; Page 72; 96pp; English.

CC This invention describes a novel nucleic acid encoding a replication
 CC protein or a plasmid stability protein. The product of the invention is
 CC useful for expression of nucleic acid such as genes encoding enzymes
 CC involved in the production of isoprenoid molecules, polyhydroxyalkanoic
 CC acid (PHA) synthases, carotenoid biosynthesis enzymes, nitrile
 CC hydratases, ethylene forming enzyme, pyruvate decarboxylase, alcohol
 CC dehydrogenase, terpene synthases, and cholesterol oxidase in an
 CC actinomycetale bacteria. The replication protein or plasmid stability
 CC protein are useful in cloning and expression vectors and particularly in
 CC shuttle vectors for the expression of homologous and heterologous genes
 CC in *Rhodococcus* sp. This sequence represents the plasmid pRHB171 DNA
 CC described in the disclosure of the invention.

XX Sequence 9652 BP; 2097 A; 2729 C; 2559 G; 2267 T; 0 other;
 SQ Query Match 100.0%; Score 1140; DB 24; Length 9652;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1140; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGACGAGCGTAAGTGTGTAACACCTTTCCGGCAAAAGACCGGCTCCCGTCTGATGCG 60
 DB 5052 ATGACGAGCGTAAGTGTGTAACACCTTTCCGGCAAAAGACCGGCTCCCGTCTGATGCG 4993

QY 61 TCCGATTAAGCGCGGCAATCCGCGACGAATCGGACCCAACTTCAACAAATCACCAGTCA 120
 DB 4992 TCCGATTAAGCGCGGCAATCCGCGACGAATCGGACCCAACTTCAACAAATCACCAGTCA 4933

QY 121 GAAACATTAAACCCCTGTGCGCCCGGATTTCTGGGTGTAACGGTGTGACATTGTCAAC 180
 DB 4932 GAAACATTAAACCCCTGTGCGCCCGGATTTCTGGGTGTAACGGTGTGACATTGTCAAC 4873

QY 181 GGTCCGAAAGGTTCTGATTCGGAAGGCTTCTTCCCTGCGGAAAGGCTGATCTGCCCC 240
 DB 4872 GGTCCGAAAGGTTCTGATTCGGAAGGCTTCTTCCCTGCGGAAAGGCTGATCTGCCCC 4813

QY 241 TGTGTGCGGGAAGGTCGATGCAATGTCGACGACGAATTTCTCAAGTTGTTGCTCAT 300
 DB 4812 TGTGTGCGGGAAGGTCGATGCAATGTCGACGACGAATTTCTCAAGTTGTTGCTCAT 4753

QY 301 CAATCGGGACTGGATCTGTTGGATGTAACGATGCAATGCGCCATACAGCTGGTCAG 360
 DB 4752 CAATCGGGACTGGATCTGTTGGATGTAACGATGCAATGCGCCATACAGCTGGTCAG 4693

QY 361 CGGCTCCAGCACTGATGATGATCTTCCGAGCTTGGAAAGTGTGACCAACGGTCTGT 420
 DB 4692 CGGCTCCAGCACTGATGATGATCTTCCGAGCTTGGAAAGTGTGACCAACGGTCTGT 4633

QY 421 CGTTGCGTACGGAACGTGAATGTACGCTGCGACGGAATGCTGCGCTGTTGAATC 480
 DB 4632 CGTTGCGTACGGAACGTGAATGTACGCTGCGACGGAATGCTGCGCTGTTGAATC 4573

QY 481 ACTACGGAAGAAACCGCTGCGACGTCACGTTACCGGCTACTCATGTTCAAGTGAAC 540
 DB 4572 ACTACGGAAGAAACCGCTGCGACGTCACGTTACCGGCTACTCATGTTCAAGTGAAC 4513

QY 541 GTGAGTGAAGACATCTCGAATCCTTCTCGGATCGAATGTTGATCGGTGACTTCCAAA 600
 DB 4512 GTGAGTGAAGACATCTCGAATCCTTCTCGGATCGAATGTTGATCGGTGACTTCCAAA 4453

QY 601 CTGCTATCTCGGATTTGCTGCGCACCTACGTAATTCGGGTGCTTCGATGTACGAAG 660
 DB 4452 CTGCTATCTCGGATTTGCTGCGCACCTACGTAATTCGGGTGCTTCGATGTACGAAG 4393

QY 661 ATCGCGGTGAAGCTGATCAAGTTCTGCTGCTGATCTGACGAAATTTGACTGCGGTT 720
 DB 4392 ATCGCGGTGAAGCTGATCAAGTTCTGCTGCTGATCTGACGAAATTTGACTGCGGTT 4333

QY 721 GGTATGAGGTTGTGATGTCGACGGAAGAAAGTGTGTCACATGCGAACCGTGCACCTGG 780
 DB 4332 GGTATGAGGTTGTGATGTCGACGGAAGAAAGTGTGTCACATGCGAACCGTGCACCTGG 4273

QY 781 GAAATCGCTGTTGATGTCAGTGGCGGGGATCCACAGGCTTGGAACTGTGGCGGAATTT 840
 DB 4272 GAAATCGCTGTTGATGTCAGTGGCGGGGATCCACAGGCTTGGAACTGTGGCGGAATTT 4213

QY 841 GAGTTGGTTGATGATGAGGACGTCGCGGCAATCGGCTGGTCCCGTGAATCGGCGGACT 900
 DB 4212 GAGTTGGTTGATGATGAGGACGTCGCGGCAATCGGCTGGTCCCGTGAATCGGCGGACT 4153

QY 901 GGTCTTGGGCGAAGACTAAGATCTGATGCTTGAACAGGAAGAACTTGCCTCCGCTC 960
 DB 4152 GGTCTTGGGCGAAGACTAAGATCTGATGCTTGAACAGGAAGAACTTGCCTCCGCTC 4093

QY 961 ATGTTGCGATCATTCGCGCGGATCGTGAATGATGATTCGACTTGTGCGCTTAACCTC 1020

DB 4092 ATGTTGCGATCATTCGCGCGGATCGTGAATGATGATTCGGAACCTTGTGCGCTTAACCTC 4033

QY 1021 TTCCGCGAAGTCTCGAATCTGTCGAACCTGCGGCACTTGGAAATCTTGTGATCAC 1080
 DB 4032 TTCCGCGAAGTCTCGAATCTGTCGAACCTGCGGCACTTGGAAATCTTGTGATCAC 3973

QY 1081 TTGCATTTATCGATTTCCCGGACGAGATGTGCGGCCCCCGGATTAATTCGTTGCAAGTCA 1140
 DB 3972 TTGCATTTATCGATTTCCCGGACGAGATGTGCGGCCCCCGGATTAATTCGTTGCAAGTCA 3913

RESULT 4
 AB076125/c
 ID AB076125 standard; DNA; 11241 BP.

AB076125;

13-JUN-2003 (first entry)

Plasmid pRHR17 DNA.

Plasmid stability protein; replication protein; ethylene forming enzyme;

carotenoid biosynthesis enzyme; isoprenoid; pyruvate decarboxylase;

polyhydroxyalkanoic acid synthase; pHA synthase; nitrile hydratase;

alcohol dehydrogenase; terpene synthase; cholesterol oxidase;

shuttle vector; circular; ds.

Synthetic.

WO200255709-A2.

18-JUL-2002.

12-DEC-2001; 2001WO-US47868.

12-DEC-2000; 2000US-254868P.

(DUPO) DU PONT DE NEMOURS & CO E. I.

Brämmuccl MG, Cheng Q, Kostichka KN, Tomb J,

WPI; 2002-557827/59.

New nucleic acid molecule encoding replication protein/plasmid

stability protein, useful in cloning and expression vectors,

particularly shuttle vectors for expression of heterologous genes in

Rhodococcus species

Claim 26; Page 71-72; 96pp; English.

This invention describes a novel nucleic acid encoding a replication

protein or a plasmid stability protein. The product of the invention is

useful for expression of nucleic acid such as genes encoding enzymes

involved in the production of isoprenoid molecules, polyhydroxyalkanoic

acid (PHA) synthases, carotenoid biosynthesis enzymes, nitrile

hydratases, ethylene forming enzyme, pyruvate decarboxylase, alcohol

dehydrogenase, terpene synthases, and cholesterol oxidase in an

Actinomycetales bacteria. The replication protein or plasmid stability

shuttle vectors are useful in cloning and expression vectors and particularly in

shuttle vectors for the expression of homologous and heterologous genes

in Rhodococcus sp. This sequence represents the plasmid pRHR17 DNA

described in the disclosure of the invention.

Sequence 11241 BP; 2517 A; 3109 C; 2945 G; 2670 T; 0 other;

Query Match 100.0%; Score 1140; DB 24; Length 11241;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 1140; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGACGAGCGTAAGTGTGTAACACCTTTCCGGCAAAAGACCGGCTCCCGTCTGATGCG 60
 DB 6641 ATGACGAGCGTAAGTGTGTAACACCTTTCCGGCAAAAGACCGGCTCCCGTCTGATGCG 6582

```

QY 61 TCCGATTAAGCGGCGCATCCGCGACGAACTCGGACCCCAACTTCAACAAATCACACGCTCA 120
DB 6581 TCCGATTAAGCGGCGCATCCGCGACGAACTCGGACCCCAACTTCAACAAATCACACGCTCA 6522
QY 121 GAAACATTTAAAGCCCTGTGCGCGCGCGATTTCTGCGGTGAACGCTGTGACCATTTGTCAAC 180
DB 6521 GAAACATTTAAAGCCCTGTGCGCGCGCGATTTCTGCGGTGAACGCTGTGACCATTTGTCAAC 6462
QY 181 GGTCCGAAAGGTTCTGGATTCGGAAGCCCTTCTCTGCGGAAAGGCGCTGATCTGCCCC 240
DB 6461 GGTCCGAAAGGTTCTGGATTCGGAAGCCCTTCTCTCTGCGGAAAGGCGCTGATCTGCCCC 6402
QY 241 TCTGTGCGGCGGAAAGTCCGTGACATCTGTGACAGAGAAATTTCTCAATTTGTGTCTAT 300
DB 6401 TCTGTGCGGCGGAAAGTCCGTGACATCTGTGACAGAGAAATTTCTCAATTTGTGTCTAT 6342
QY 301 CAACCTCGGAGCTGGATCTGTGGATGTGACGATACATGACATGCGGCAATCAGCTGCTGAC 360
DB 6341 CAACCTCGGAGCTGGATCTGTGGATGTGACGATACATGACATGCGGCAATCAGCTGCTGAC 6282
QY 361 CGGCTCCACGACCTATGACCTGGAATTTTCGCGAGCCCTGGAAAGCTGCGACCAACGCTGCT 420
DB 6281 CGGCTCCACGACCTATGACCTGGAATTTTCGCGAGCCCTGGAAAGCTGCGACCAACGCTGCT 6222
QY 421 CGTTGGCGTACGGAACGTGAATGTACGCTGCGACGAGATACGCTGCGCGCTGTGAAATTC 480
DB 6221 CGTTGGCGTACGGAACGTGAATGTACGCTGCGACGAGATACGCTGCGCGCTGTGAAATTC 6162
QY 481 ACTCAGGAGAAAAAGGCTGGCGACGCTCCAGCTCAGCGCTACTCATGTTCACTAGTGTGAC 540
DB 6161 ACTCAGGAGAAAAAGGCTGGCGACGCTCCAGCTCAGCGCTACTCATGTTCACTAGTGTGAC 6102
QY 541 GTGAGTGAGAACATCTCTGCAATCCTTCTCGGATGCGATGTCGATCGTGTGACCTTCCAA 600
DB 6101 GTGAGTGAGAACATCTCTGCAATCCTTCTCGGATGCGATGTCGATGTCGATGTCGATGTC 6042
QY 601 CTGATCTCTGCGGATTTCTGCGCGACATCACTGATTTGCGGTGCTCGATGTGACAAAG 660
DB 6041 CTGATCTCTGCGGATTTCTGCGCGACATCACTGATTTGCGGTGCTCGATGTGACAAAG 5982
QY 661 ATCCGCGGTGAGAGCTGATCAAGTTCGCTGCGTATCTGACGAAATTCGATCTGCGCT 720
DB 5981 ATCCGCGGTGAGAGCTGATCAAGTTCGCTGCGTATCTGACGAAATTCGATCTGCGCT 5922
QY 721 GGTATGAGAGTTGTTGTTGCGCGACGAGAAAAAGTGTGACATGCGACCGTGTGACCTCTG 780
DB 5921 GGTATGAGAGTTGTTGTTGCGCGACGAGAAAAAGTGTGACATGCGACCGTGTGACCTCTG 5862
QY 781 GAAATGCTGTTGATGCACTGCGCGGAGATCCAAAGCTTGGAACTGTGCGGACGAAATTT 840
DB 5861 GAAATGCTGTTGATGCACTGCGCGGAGATCCAAAGCTTGGAACTGTGCGGACGAAATTT 5802
QY 841 GATTTGTTGATGCACTGCGCGGAGATCCAAAGCTTGGAACTGTGCGGACGAAATTT 900
DB 5801 GATTTGTTGATGCACTGCGCGGAGATCCAAAGCTTGGAACTGTGCGGACGAAATTT 5742
QY 901 GGTCTTGGGCGCAGAACCTAACAGATGCTCAGATGTTGAGCAGAGAAATTCGCGCGCTC 960
DB 5741 GGTCTTGGGCGCAGAACCTAACAGATGCTCAGATGTTGAGCAGAGAAATTCGCGCGCTC 5682
QY 961 ATGTTGCGATCTTCGCGCGGCGATCGTGTGATATGATTCGCACTTGTGCGCCTTACGCT 1020
DB 5681 ATGTTGCGATCTTCGCGCGGCGATCGTGTGATATGATTCGCACTTGTGCGCCTTACGCT 5622
QY 1021 TTGCGGAGAGATCTCTCGGACCTGTCGAAAGCTGCGCGACCTTGGGAAATTCCTGCTGATCAC 1080
DB 5621 TTGCGGAGAGATCTCTCGGACCTGTCGAAAGCTGCGCGACCTTGGGAAATTCCTGCTGATCAC 5562
QY 1081 TTGATTTATGATTTGCGCGGAGATGTCGCGCGCCCGATATATCGTTTCCGCAAGTGA 1140
DB 5561 TTGATTTATGATTTGCGCGGAGATGTCGCGCGCCCGATATATCGTTTCCGCAAGTGA 5502

```

```

RESULT 5
ABQ76138/c
ID ABQ76138 standard; DNA; 20 BP.
XX
XX
AC ABQ76138;
XX
XX
DT 13-JAN-2003 (first entry)
XX
DE Rhodococcus AN12 derived plasmid pAN12 div PCR primer Rep1.
XX
KW Plasmid stability protein; replication protein; ethylene forming enzyme;
KW carotenoid biosynthesis enzyme; isoprenoid; pyruvate decarboxylase;
KW polyhydroxyalkanoic acid synthase; PRA synthase; nitrile hydratase;
KW alcohol dehydrogenase; terpene synthase; cholesterol oxidase;
KW shuttle vector; div; PCR; primer; ss.
XX
OS Rhodococcus erythropolis.
OS Synthetic.
XX
PN WO200255709-A2.
XX
PD 18-JUL-2002.
XX
PF 12-DEC-2001; 2001WO-US47868.
XX
PR 12-DEC-2000; 2000US-254868P.
XX
PA (DUPO ) DU PONT DE NEMOURS & CO E I.
PI Bramucci MG, Cheng Q, Kostichka KN, Tomb J;
PI WPI; 2002-557827/59.
DR
XX
XX
PT New nucleic acid molecule encoding replication protein/plasmid
PT stability protein, useful in cloning and expression vectors,
PT particularly shuttle vectors for expression of heterologous genes in
PT Rhodococcus species -
XX
XX
PS Example 7; Page 47; 96bp; English.
XX
CC This invention describes a novel nucleic acid encoding a replication
CC protein or a plasmid stability protein. The product of the invention is
CC useful for expression of nucleic acid such as genes encoding enzymes
CC involved in the production of isoprenoid molecules, polyhydroxyalkanoic
CC acid (PHA) synthases, carotenoid biosynthesis enzymes, nitrile
CC hydratases, ethylene forming enzyme, pyruvate decarboxylase, alcohol
CC dehydrogenase, terpene synthases, and cholesterol oxidase in an
CC Actinomycetales bacteria. The replication protein or plasmid stability
CC protein are useful in cloning and expression vectors and particularly in
CC shuttle vectors for the expression of homologous and heterologous genes
CC in Rhodococcus sp. This sequence represents the Rhodococcus AN12 derived
CC plasmid pAN12 div PCR primer Rep1 described in the disclosure of the
CC invention.
XX
SQ Sequence 20 BP; 6 A; 5 C; 3 G; 6 T; 0 other;
XX
XX
Query Match 1.8%; Score 20; DB 24; Length 20;
Best Local Similarity 100.0%; Pred. No. 7.4;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1119 GATTAATATCGTTCCGCAAGT 1138
DB 20 GATTAATATCGTTCCGCAAGT 1
XX
XX
RESULT 6
ABLS7927/c
ID ABLS7927 standard; cDNA; 167 BP.
XX
XX
AC ABLS7927;
XX
XX
DT 22-JUL-2002 (first entry)
XX

```

```

DE Human VG51 exon 8.
XX
XX Antiafasmatic; anxiolytic; antiepileptic; antihypertensive; human;
KW psychotropic; glutamate transporter; transporter; GABA;
KW gamma-aminobutyric acid transporter; GABA transporter; neurotransmitter;
KW asthma; anxiety; epilepsy; hypertension; psychiatric disorder;
KW neurotic disorder; VG51; ss.
XX
OS Homo sapiens.
XX
XX WO200071709-A1.
XX
XX 30-NOV-2000.
XX
XX 19-MAY-2000; 2000WO-FR01383.
XX
XX 21-MAY-1999; 99FR-0006525.
XX
XX (INRM ) INSERM INST NAT SANTE & RECH MEDICALE.
XX
XX Giros B, Gasnier B, Sagne C, El Mestikawy S, Hamon M;
XX WPI; 2001-025160/03.
XX
XX New mammalian amino acid transporter, used e.g. to screen for
XX psychotropic agents, is high capacity but low affinity transporter of
XX gamma-aminobutyric acid -
XX
XX Claim 5; Page 96; 103pp; French.
XX
XX The present sequence is exon 8 of human VG51 cDNA sequence, a glutamate/
XX gamma-aminobutyric acid (GABA) transporter. GABA and glutamate are
XX neurotransmitters. The transporter can be used to produce specific
XX antibodies to screen for binding agents. Modulators of the transporter
XX are useful for treating disorders associated with deregulated
XX glutamate/GABA transport, e.g. asthma, anxiety, epilepsy, hypertension
XX and other psychiatric and neurotic disorders, while determining levels of
XX the transporter and its coding sequence can be used for diagnosis of such
XX disorders. The full length coding sequence is given in ABR57930.
XX
XX
SQ Sequence 167 BP; 40 A; 51 C; 37 G; 39 T; 0 other;
XX
XX Query Match 1.8%; Score 20; DB 23; Length 167;
XX Best Local Similarity 100.0%; Pred. No. 6.7;
XX Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 324 GATGTGACGATGACCATGC 343
XX |||||
XX Db 81 GATGTGACGATGACCATGC 62
XX
XX RESULT 7
XX AAS90653/c
XX ID AAS90653 standard; cDNA; 739 BP.
XX
XX AC AAS90653;
XX
XX DT 13-FEB-2002 (first entry)
XX
XX DE DNA encoding novel human diagnostic protein #26457.
XX
XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
XX OS Homo sapiens.
XX
XX PN WO200175067-A2.
XX
XX PD 11-OCT-2001.
XX
XX PF 30-MAR-2001; 2001WO-US08631.
XX
XX PR 31-MAR-2000; 2000US-0540217.

```

```

PR 23-AUG-2000; 2000US-0649167.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Drmanac RT, Liu C, Tang YT;
XX WPI; 2001-639362/73.
XX
XX P-PSDB; ABG26466.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
XX diagnostics, forensics, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits and to assess
XX biodiversity -
XX
XX Claim 1; SEQ ID No 26457; 103pp; English.
XX
XX The invention relates to isolated polynucleotide (I) and
XX polypeptide (II) sequences. (I) is useful as hybridisation probes,
XX polymerase chain reaction (PCR) primers, oligomers, and for chromosome
XX and gene mapping, and in recombinant production of (II). The
XX polynucleotides are also used in diagnostics as expressed sequence tags
XX for identifying expressed genes. (I) is useful in gene therapy techniques
XX to restore normal activity of (II) or to treat disease states involving
XX (II). (II) is useful for generating antibodies against it, detecting or
XX quantitating a polypeptide in tissue; as molecular weight markers and as
XX a food supplement. (II) and its binding partners are useful in medical
XX imaging of sites expressing (II). (I) and (II) are useful for treating
XX disorders involving aberrant protein expression or biological activity.
XX The polypeptide and polynucleotide sequences have applications in
XX diagnostics, forensics, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits to assess biodiversity
XX and to produce other types of data and products dependent on DNA and
XX amino acid sequences. AAS64197-AAS94564 represent novel human
XX diagnostic coding sequences of the invention.
XX Note: The sequence data for this patent did not appear in the printed
XX specification, but was obtained in electronic format directly from WIPO
XX atftp.wipo.int/pub/published_pct_sequences.
XX
XX
SQ Sequence 739 BP; 145 A; 231 C; 168 G; 193 T; 2 other;
XX
XX Query Match 1.8%; Score 20; DB 23; Length 739;
XX Best Local Similarity 100.0%; Pred. No. 6.2;
XX Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 324 GATGTGACGATGACCATGC 343
XX |||||
XX Db 470 GATGTGACGATGACCATGC 451
XX
XX RESULT 8
XX ABV23244/c
XX ID ABV23244 standard; cDNA; 947 BP.
XX
XX AC ABV23244;
XX
XX DT 16-SEP-2002 (first entry)
XX
XX DE Human prostate expression marker cDNA 23235.
XX
XX KW Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;
XX KW pharmacogenomic marker; gene; ss.
XX
XX OS Homo sapiens.
XX
XX PN WO200160860-A2.
XX
XX PD 23-AUG-2001.
XX
XX PF 20-FEB-2001; 2001WO-US05171.
XX
XX PR 17-FEB-2000; 2000US-183319P.
XX
XX PR 16-MAR-2000; 2000US-189862P.
XX
XX PR 25-MAY-2000; 2000US-207454P.

```

PR 09-JUN-2000; 2000US-211314P.
 PR 18-JUL-2000; 2000US-219007P.
 PR 13-DEC-2000; 2000US-255281P.
 XX
 PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
 XX
 PI Schlegel R, Endege WO, Monahan JE;
 XX
 DR WPI, 2001-662795/76.
 XX
 PT Novel isolated nucleic acid molecule associated with cancerous state of
 PT prostate cells and correlating with presence of prostate cancer, useful
 PT for detecting presence of prostate cancer, stage of prostate cancer -
 XX
 PS Claim 1; Page 4190; 11750pp; English.
 XX
 CC The invention relates to an isolated nucleic acid molecule (I) comprising
 CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
 CC specification or its complement. (I) is useful for:
 CC (a) assessing whether a patient is afflicted with prostate cancer;
 CC (b) monitoring the progression of prostate cancer in a patient;
 CC (c) assessing the efficacy of a test compound to inhibit prostate
 CC cancer in a patient;
 CC (d) assessing the efficacy of a therapy for inhibiting prostate cancer
 CC in a patient;
 CC (e) selecting a composition for inhibiting prostate cancer in a patient;
 CC (f) assessing the prostate cell carcinogenic potential of a compound;
 CC (g) determining whether prostate cancer has metastasized in a patient;
 CC (h) assessing the aggressiveness or indolence of prostate cancer in a
 CC patient;
 CC (I) is also useful as a pharmacodynamic or pharmacogenomic marker.
 XX
 SQ Sequence 947 BP; 172 A; 307 C; 223 G; 241 T; 4 other;
 XX
 Query Match 1.8%; Score 20; DB 23; Length 947;
 Best Local Similarity 100.0%; Pred. No. 6.1;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 324 GATGTGACGATGACCATGC 343
 Db 470 GATGTGACGATGACCATGC 451
 XX
 RESULT 9
 ABV29088/C
 ID ABV29088 standard; cDNA; 947 BP.
 XX
 AC ABV29088;
 XX
 DT 16-SEP-2002 (first entry)
 XX
 DE Human prostate expression marker cDNA 29079.
 XX
 KM Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;
 KM pharmacogenomic marker; gene; ss.
 XX
 OS Homo sapiens.
 OS
 XX
 PN WO200160860-A2.
 XX
 PD 23-AUG-2001.
 XX
 PF 20-FEB-2001; 2001WO-US05171.
 XX
 PR 17-FEB-2000; 2000US-183319P.
 PR 16-MAR-2000; 2000US-189622P.
 PR 25-MAY-2000; 2000US-207454P.
 PR 09-JUN-2000; 2000US-211314P.
 PR 18-JUL-2000; 2000US-219007P.
 PR 13-DEC-2000; 2000US-255281P.
 XX
 PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
 XX

PI Schlegel R, Endege WO, Monahan JE;
 XX
 DR WPI, 2001-662795/76.
 XX
 PT Novel isolated nucleic acid molecule associated with cancerous state of
 PT prostate cells and correlating with presence of prostate cancer, useful
 PT for detecting presence of prostate cancer, stage of prostate cancer -
 XX
 PS Claim 1; Page 6170-6171; 11750pp; English.
 XX
 CC The invention relates to an isolated nucleic acid molecule (I) comprising
 CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
 CC specification or its complement. (I) is useful for:
 CC (a) assessing whether a patient is afflicted with prostate cancer;
 CC (b) monitoring the progression of prostate cancer in a patient;
 CC (c) assessing the efficacy of a test compound to inhibit prostate
 CC cancer in a patient;
 CC (d) assessing the efficacy of a therapy for inhibiting prostate cancer
 CC in a patient;
 CC (e) selecting a composition for inhibiting prostate cancer in a patient;
 CC (f) assessing the prostate cell carcinogenic potential of a compound;
 CC (g) determining whether prostate cancer has metastasized in a patient;
 CC (h) assessing the aggressiveness or indolence of prostate cancer in a
 CC patient;
 CC (I) is also useful as a pharmacodynamic or pharmacogenomic marker.
 XX
 SQ Sequence 947 BP; 172 A; 307 C; 223 G; 241 T; 4 other;
 XX
 Query Match 1.8%; Score 20; DB 23; Length 947;
 Best Local Similarity 100.0%; Pred. No. 6.1;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 324 GATGTGACGATGACCATGC 343
 Db 470 GATGTGACGATGACCATGC 451
 XX
 RESULT 10
 ABL57930/C
 ID ABL57930 standard; cDNA; 1428 BP.
 XX
 AC ABL57930;
 XX
 DT 22-JUL-2002 (first entry)
 XX
 DE Human VG51 coding sequence.
 XX
 KM Antiasmatic; anxiolytic; antiepileptic; antihypertensive; human;
 KM psychotropic; glutamate transporter; transporter; GABA;
 KM gamma-aminobutyric acid transporter; GABA transporter; neurotransmitter;
 KM asthma; anxiety; epilepsy; hypertension; psychiatric disorder;
 KM neurotic disorder; VG51; gene; ss.
 XX
 OS Homo sapiens.
 OS
 XX
 FH Key Location/Qualifiers
 FH CDS 1..1428
 FT /*tag= a
 FT /partial
 FT /product= "VG51"
 FT /note= "No stop codon given"
 XX
 PN WO200071709-A1.
 XX
 PD 30-NOV-2000.
 XX
 PR 19-MAY-2000; 2000WO-FR01383.
 XX
 PR 21-MAY-1999; 99FR-0006525.
 XX
 PA (INRM) INSERM INST NAT SANTE & RECH MEDICALE.
 XX
 PI Giros B, Gasnier B, Sagne C, El Mestikawy S, Hamon M;
 XX

XX WP1: 2001-025160/03.
 DR P-PSDB; ABB76941.
 XX
 PT New mammalian amino acid transporter, used e.g. to screen for
 PT psychotropic agents, is high capacity but low affinity transporter of
 PT gamma-aminobutyric acid
 XX
 PS Claim 5; Fig 13; 103pp; French.
 XX
 CC The present sequence is the coding sequence for human VGLT, a glutamate/
 CC gamma-aminobutyric acid (GABA) transporter. GABA and glutamate are
 CC neurotransmitters. The transporter can be used to produce specific
 CC antibodies, to screen for binding agents. Modulators of the transporter
 CC are useful for treating disorders associated with deregulated
 CC glutamate/GABA transport, e.g. asthma, anxiety, epilepsy, hypertension
 CC and other psychiatric and neurotic disorders, while determining levels of
 CC the transporter and its coding sequence can be used for diagnosis of such
 CC disorders.
 XX
 SQ Sequence 1428 BP; 289 A; 435 C; 337 G; 367 T; 0 other;
 Query Match 1.8%; Score 20; DB 23; Length 1428;
 Best Local Similarity 100.0%; Pred. No. 6;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 324 GATGGTGACGATGACCATGC 343
 Db 903 GATGGTGACGATGACCATGC 884
 RESULT 11
 ABX70643/C
 ID ABX70643 standard; cDNA; 1501 BP.
 AC ABX70643;
 XX
 DT 04-MAR-2003 (first entry)
 XX
 DE Human cDNA encoding NOV7b.
 XX
 KW Human; ss; gene; NOVX; metabolic disorder; diabetes; cardiomyopathy;
 KW obesity; infectious disease; anorexia; neurodegenerative disorder;
 KW Alzheimer's disease; Parkinson's disease; immune disorder;
 KW haematopoietic disorder; dyslipidaemia; metabolic disturbance;
 KW metabolic syndrome X; wasting disorder; cancer; gene therapy.
 XX
 OS Homo sapiens.
 XX
 PN WO200281518-A2.
 XX
 PD 17-OCT-2002.
 XX
 PF 21-FEB-2002; 2002WO-US05374.
 XX
 PR 21-FEB-2001; 2001US-270220P.
 PR 21-FEB-2001; 2001US-270523P.
 PR 23-FEB-2001; 2001US-270797P.
 PR 23-FEB-2001; 2001US-270810P.
 PR 08-MAR-2001; 2001US-274295P.
 PR 16-MAR-2001; 2001US-276400P.
 PR 16-MAR-2001; 2001US-276677P.
 PR 26-MAR-2001; 2001US-278796P.
 PR 04-APR-2001; 2001US-281521P.
 PR 25-APR-2001; 2001US-286548P.
 PR 13-AUG-2001; 2001US-311980P.
 PR 10-SEP-2001; 2001US-318526P.
 PR 17-SEP-2001; 2001US-322712P.
 PR 18-OCT-2001; 2001US-330307P.
 XX
 PA (CUTRA-) CUBAGEN CORP.
 XX
 PI Pena CEA, Shinkets RA, Li L, Shenoy SG, Kekuda R, Spytek KA;

PI Vernet CM, Malyankar UL, Guo X, Gusev VY, Casman ST, Boldog FL;
 PI Furtak K, Tohernev VT, Patturajan M, Gangoli EA, Padigaru M;
 PI Liu X, Baumgartner JC, Gerlach VL, Spaderma SK, Zernusen BD;
 XX
 DR WP1: 2003-046859/04.
 DR P-PSDB; ABUS2588.
 XX
 PT New isolated NOVX polypeptide useful for treating cardiomyopathy,
 PT atherosclerosis, metabolic disorders, diabetes, obesity, infectious
 PT disease, anorexia, neurodegenerative disorders, Alzheimer's disease and
 PT cancer
 XX
 PS Claim 9; Page 57; 479pp; English.
 XX
 CC The invention relates to an isolated polypeptide termed NOVX (NOV1,
 CC 2a, 2b, 3a, 3b, 4a, 4b, 5, 6, 7a-c, 8a-e, 9a-b, 10, 11, 12a-c, 13, 14,
 CC 15, 16a-d, 17a-b, 18, 19, 20a-b, 21-30) appearing as ABUS2578-ABUS2624),
 CC a variant of NOVX, a mature form of NOVX, and a variant of the mature
 CC form of NOVX. Also included are a nucleic acid molecule (NOVX NA)
 CC encoding NOVX, or a fragment or complement of NOVX NA, a vector
 CC comprising NOVX NA, a cell comprising the vector, an anti-NOVX antibody
 CC (ab), determining the presence or amount of NOVX or NOVX NA in a sample,
 CC and identifying an agent that binds or modulates the expression or
 CC activity of NOVX. NOVX NA or ab is useful for treating or
 CC preventing a NOVX-associated disorder in a subject, preferably human. Ab
 CC is useful for determining the presence or amount of NOVX in a sample.
 CC NOVX is useful for identifying an agent that binds to NOVX. NOVX
 CC NA or ab is useful for treating metabolic disorders, diabetes,
 CC cardiomyopathy, obesity, infectious disease, anorexia, neurodegenerative
 CC disorders, Alzheimer's disease, Parkinson's disease, immune disorders,
 CC haematopoietic disorders, and various dyslipidaemias, metabolic
 CC disturbances associated with obesity, the metabolic syndrome X and
 CC wasting disorders associated with chronic diseases, various cancers,
 CC endocrine, connective tissue, blood, vascular, skin, renal, bone, brain,
 CC muscle disorders, or bacterial, fungal, protozoal or viral infections.
 CC NOVX, NOVX NA or ab is useful in screening assays, detection assays,
 CC predictive medicine, and in methods of treatment. NOVX is useful as
 CC immunogen, to screen for potential ant/agonist compounds, and as bait
 CC protein in a two-hybrid or three-hybrid assay. NOVX NA is useful in gene
 CC therapy, to express NOVX, to detect NOVX mRNA or a genetic lesion in a
 CC NOVX gene, and to modulate NOVX activity. The cell is useful for
 CC producing non-human transgenic animals. Ab is useful for isolating, and
 CC purifying NOVX and to monitor protein levels in tissue as part of a
 CC clinical testing procedure. The present sequence encodes a NOVX protein.
 XX
 SQ Sequence 1501 BP; 303 A; 459 C; 361 G; 378 T; 0 other;
 Query Match 1.8%; Score 20; DB 25; Length 1501;
 Best Local Similarity 100.0%; Pred. No. 6;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 324 GATGGTGACGATGACCATGC 343
 Db 964 GATGGTGACGATGACCATGC 945
 RESULT 12
 ABX70642/C
 ID ABX70642 standard; cDNA; 1513 BP.
 AC ABX70642;
 XX
 DT 04-MAR-2003 (first entry)
 XX
 DE Human cDNA encoding NOV7a.
 XX
 KW Human; ss; gene; NOVX; metabolic disorder; diabetes; cardiomyopathy;
 KW obesity; infectious disease; anorexia; neurodegenerative disorder;
 KW Alzheimer's disease; Parkinson's disease; immune disorder;
 KW haematopoietic disorder; dyslipidaemia; metabolic disturbance;
 KW metabolic syndrome X; wasting disorder; cancer; gene therapy; SNP;
 KW single nucleotide polymorphism.

OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT variation replace (256,G)
FT /*tag= a
FT /standard_name= "Single nucleotide polymorphism"
FT replace (519,T)
FT /*tag= b
FT /standard_name= "Single nucleotide polymorphism"
FT replace (1153,T)
FT /*tag= c
FT /standard_name= "Single nucleotide polymorphism"
XX
XX WO200281518-A2.
XX
XX 17-OCT-2002.
XX
XX 21-FEB-2002; 2002WO-US05374.
XX
XX 21-FEB-2001; 2001US-270220P.
XX 21-FEB-2001; 2001US-270523P.
XX 23-FEB-2001; 2001US-270797P.
XX 23-FEB-2001; 2001US-270810P.
XX 08-MAR-2001; 2001US-274295P.
XX 16-MAR-2001; 2001US-276400P.
XX 16-MAR-2001; 2001US-276677P.
XX 26-MAR-2001; 2001US-278796P.
XX 04-APR-2001; 2001US-281521P.
XX 25-APR-2001; 2001US-286548P.
XX 13-AUG-2001; 2001US-311980P.
XX 10-SEP-2001; 2001US-318526P.
XX 17-SEP-2001; 2001US-322712P.
XX 18-OCT-2001; 2001US-330307P.
XX
XX (CURA-) CURAGEN CORP.
XX
XX Pena CEA, Shinkets RA, Li L, Shenoy SG, Kekuda R, Spytek KA,
XX Vermet CM, Malyankar UL, Guo X, Gusev VY, Casman SJ, Boldog FL,
XX Fortak K, Tcherenev VY, Patcurajan M, Gangoli EA, Padigaru M,
XX Liu X, Baumgartner UC, Gerlach VL, Spaderina SK, Zethusen BD;
XX WPI; 2003-046859/04.
XX P-PSDB; ABUS2587.
XX
XX New isolated NOVX polypeptide useful for treating cardiomyopathy,
XX PT atherosclerosis, metabolic disorders, diabetes, obesity, infectious
XX PT disease, anorexia, neurodegenerative disorders, Alzheimer's disease and
XX PT cancer -
XX
XX Claim 9; Page 55; 479pp; English.
XX
XX The invention relates to an isolated polypeptide termed NOVX (NOV1,
XX 2a, 2b, 3a, 3b, 4a, 4b, 5, 6, 7a-c, 8a-e, 9a-b, 10, 11, 12a-c, 13, 14,
XX 15, 16a-d, 17a-b, 18, 19, 20a-b, 21-30) appearing as ABUS2578-ABUS2624),
XX a variant of NOVX, a mature form of NOVX, and a variant of the mature
XX form of NOVX. Also included are a nucleic acid molecule (NOVX NA)
XX encoding NOVX, or a fragment or complement of NOVX NA, a vector
XX comprising NOVX NA, a cell comprising the vector, an anti-NOVX antibody
XX (ab), determining the presence or amount of NOVX or NOVX NA in a sample,
XX and identifying an agent that binds or modulates the expression or
XX activity of NOVX. NOVX, NOVX NA or ab is useful for treating or
XX preventing a NOVX-associated disorder in a subject, preferably human. Ab
XX is useful for determining the presence or amount of NOVX in a sample.
XX NOVX is useful for identifying an agent that binds to NOVX. NOVX
XX NA or ab is useful for treating metabolic disorders, diabetes,
XX cardiomyopathy, obesity, infectious disease, anorexia, neurodegenerative
XX disorders, Alzheimer's disease, Parkinson's disease, immune disorders,
XX haematopoietic disorders, and various dyslipidaemias, metabolic
XX disturbances associated with obesity, the metabolic syndrome X and
XX wasting disorders associated with chronic diseases, various cancers,
XX endocrine, connective tissue, blood, vascular, skin, renal, bone, brain,
XX muscle disorders, or bacterial, fungal, protozoal or viral infections.
XX NOVX, NOVX NA or ab is useful in screening assays, detection assays,

CC predictive medicine, and in methods of treatment. NOVX is useful as
CC immunogen, to screen for potential ant/agonist compounds, and as bait
CC protein in a two-hybrid or three-hybrid assay. NOVX NA is useful in gene
CC therapy, to express NOVX, to detect NOVX mRNA or a genetic lesion in a
CC NOVX gene, and to modulate NOVX activity. The cell is useful for
CC producing non-human transgenic animals. Ab is useful for isolating, and
CC purifying NOVX and to monitor protein levels in tissue as part of a
CC clinical testing procedure. The present sequence encodes a NOVX protein.
XX
SQ Sequence 1513 BP; 304 A; 462 C; 362 G; 385 T; 0 other;
XX
Query Match 1.8%; Score 20; DB 25; Length 1513;
Best Local Similarity 100.0%; Pred. No. 6;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
QY 324 GATGCTGACGATGACCATGC 343
Db 988 GATGCTGACGATGACCATGC 969
XX
RESULT 13
ID ABX70783 standard; cDNA; 1597 BP.
XX
AC ABX70783;
XX
DT 05-MAR-2003 (first entry)
XX
XX Novel human cDNA sequence #8.
XX
XX Human; gene; ss; nervous system disorder; peripheral neuropathy;
XX Huntington's disease; amyotrophic lateral sclerosis; haemophilia;
XX neurodegenerative disease; Parkinson's disease; Alzheimer's disease;
XX autoimmune disease; systemic lupus erythematosus; rheumatoid arthritis;
XX insulin-dependent diabetes mellitus; anaemia; thrombocytopaenia; wound;
XX ulcer; burn; bone disorder; osteoporosis; osteoarthritis; stroke;
XX fibrosis; reperfusion injury; infection; allergic rhinitis; asthma;
XX coagulation disorder; cancer; tumour; inflammatory disease;
XX septic shock; Crohn's disease; anaphylaxis; proliferation; chemotactic;
XX differentiation; stem cell growth factor; haematopoiesis; chemokinetic;
XX haemostatic; antiinflammatory; expressed sequence tag; EST.
XX
XX Homo sapiens.
XX
XX WO200281731-A2.
XX
XX 17-OCT-2002.
XX
XX 29-JAN-2002; 2002WO-US01222.
XX
XX 30-JAN-2001; 2001US-0774528.
XX
XX (HYSE-) HYSE INC.
XX (GOOD/) GOODRICH R W.
XX
XX Tang TY, Liu C, Zhou P, Asundi V, Zhang J, Zhao QA, Ren F;
XX Xue AJ, Yang Y, Weinman T, Wang J, Wang D, Drmanac RT;
XX WPI; 2003-058563/05.
XX
XX Novel polypeptide useful for treating neurodegenerative diseases,
XX PT myeloid or lymphoid cell disorders, bone disorders, mechanical and
XX PT traumatic disorders, coagulation disorders, and inflammatory diseases
XX -
XX
XX Claim 1; Page -; 612pp; English.
XX
XX This invention relates to the cDNA sequences encoding an isolated
XX novel human polypeptide. The protein encoded by the nucleic acid of
XX the invention is useful for treating central and peripheral nervous
XX system diseases (e.g. peripheral neuropathy, Huntington's disease,
XX amyotrophic lateral sclerosis); neurodegenerative diseases (e.g.
XX Parkinson's disease, Alzheimer's disease); autoimmune disease (e.g.

CC systemic lupus erythematosus, rheumatoid arthritis, insulin-dependent
CC diabetes mellitus); myeloid or lymphoid cell disorders (e.g. anaemia
CC and thrombocytopenia); wounds, ulcers, burns; bone disorders (e.g.
CC osteoporosis, osteoarthritis); mechanical and traumatic disorders (e.g.
CC stroke, head trauma); lung or liver fibrosis; reperfusion injury in
CC various tissues; bacterial, viral or fungal infections; allergic
CC conditions such as allergic rhinitis, asthma, coagulation disorders
CC (e.g. haemophilia); cancer and tumours; and inflammatory diseases (e.g.
CC septic shock, Crohn's disease, anaphylaxis). The protein may be used to
CC inhibit the growth, infection or function of infectious agents such as
CC bacteria, fungi, viruses, or to effect bodily characteristics,
CC biohythms or circadian cycles of rhythms. The protein may also
CC have proliferation/differentiation, stem cell growth factor,
CC haematopoiesis regulation, immune stimulating or suppressing,
CC chemotactic/chemokinetic, hemostatic and thrombolytic, receptor/ligand,
CC and antiinflammatory activities. The cDNA sequences of the invention are
CC useful for expressing recombinant protein for analysis. The present
CC sequence represents a novel human cDNA sequence of the invention,
CC this sequence is an expressed sequence tag (EST) and was identified
CC using subtractive hybridisation.

XX SQ Sequence 1597 BP, 327 A, 483 C, 383 G, 404 T, 0 other;

Query Match 1.8%; Score 20; DB 25; Length 1597;
Best Local Similarity 100.0%; Pred. No. 6;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 324 GATGTGACGATGACCATGC 343
DB 1005 GATGTGACGATGACCATGC 986

RESULT 14

AA590655
ID AA590655 standard; cDNA; 1771 BP.

XX AC AA590655;

XX DT 13-FEB-2002 (first entry)

DE DNA encoding novel human diagnostic protein #26459.

KM Human; chromosome mapping; gene mapping; gene therapy; forensic;

KW food supplement; medical imaging; diagnostic; genetic disorder; ss.

XX OS Homo sapiens.

XX PN WO200175067-A2.

XX PD 11-OCT-2001.

XX PF 30-MAR-2001; 2001WO-US08631.

XX PR 31-MAR-2000; 2000US-0540217.

XX PR 23-AUG-2000; 2000US-0649167.

XX PA (HYSB-) HYSBQ INC.

XX PI Dmanac RT, Liu C, Tang YT;

XX DR WPI; 2001-639362/73.

XX DR P-PSDB; ABG26468.

XX PT New isolated polynucleotide and encoded polypeptides, useful in
XX PT diagnostics, forensics, gene mapping, identification of mutations
XX PT responsible for genetic disorders or other traits and to assess
XX PT biodiversity -

XX PS Claim 1; SEQ ID No 26459; 103bp; English.

XX CC The invention relates to isolated polynucleotide (I) and
XX CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
XX CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome

CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AA564197-AA594564 represent novel human
CC diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

XX SQ Sequence 1771 BP, 422 A, 444 C, 529 G, 376 T, 0 other;

Query Match 1.8%; Score 20; DB 23; Length 1771;
Best Local Similarity 100.0%; Pred. No. 6;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 324 GATGTGACGATGACCATGC 343
DB 1347 GATGTGACGATGACCATGC 1366

RESULT 15
AB280236/C
ID AB280236 standard; cDNA; 2057 BP.

XX AC AB280236;

XX DT 02-JUN-2003 (first entry)

DE Human transdormin 3 encoding cDNA SEQ ID NO:5.

KM Neuroprotective; nocrotropic; cerebroprotective; analgesic; gene therapy;

KW central nervous system disorder; CNS disorder; multiple sclerosis;

KW nerve injury; neuropathic pain; stroke; trauma; non-CNS disorder; transdormin; human; transdormin 3; chromosome 5; gene; ss.

XX OS Homo sapiens.

XX FH Key Location/Qualifiers

XX FT CDS 152..1582 /*tag= a

XX FT /product= "transdormin 3"

XX FT /transl_except= (pos:1217..1219,aa:11e)

XX PN WO2003016502-A2.

XX PD 27-FEB-2003.

XX PF 21-AUG-2002; 2002WO-US26637.

XX PR 21-ANG-2001; 2001US-313907P.

XX PR 21-AUG-2002; 2002US-0225810.

XX PA (MCLA-) MCLAUGHLIN RES INST.

XX PI Birmingham JR;

XX DR WPI; 2003-278567/27.

XX DR P-PSDB; ABP96441.
XX PT New nucleic acid sequence encoding transdormin, e.g. mouse strand 1,
XX PT mouse strand 2, mouse strand 3, human strand 1, human strand 2, human strand
XX PT 3 or rat strand 1, useful for treating CNS, e.g. stroke, multiple
XX PT sclerosis, trauma, neuropathic pain -

XX
PS Claim 1; Fig 21; 177pp; English.
XX
CC The present invention describes an isolated nucleic acid sequence
CC comprising a cDNA sequence encoding mouse tramdorin (trandl 2, mouse
CC trand 3, human trand 1, human trand 2, human trand 3 or rat trand 1, or
CC the genomic sequence of mouse trand 1 or mouse trand 3. Mouse trand 1 is
CC located to chromosome 11, whereas human trand 1 is located to chromosome
CC 5q31-33. The trand sequences have neuroprotective, nootropic, analgesic
CC and cerebroprotective activities, and can be used in gene therapy. The
CC nucleic acid sequences are useful for diagnosing and treating central
CC nervous system (CNS) disorders such as multiple sclerosis, nerve injury,
CC neuropathic pain, stroke or trauma, and non-CNS disorders. The present
CC sequence encodes human trand 3, which is given in the exemplification of
CC the present invention.
XX

SQ Sequence 2057 BP; 411 A; 638 C; 503 G; 505 T; 0 other;

Query Match 1.8%; Score 20; DB 25; Length 2057;
Best Local Similarity 100.0%; Pred.No. 5.9;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 324 GATGTTGACGATGACCATGC 343
DB 1054 GATGTTGACGATGACCATGC 1035

Search completed: November 8, 2003, 21:40:10
Job time : 283 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 8, 2003, 21:30:02 ; Search time 70 Seconds
(without alignments)

7188.240 Million cell updates/sec

Title: US-10-007-527a-1

Perfect score: 1140

Sequence: 1 atgaccagcgaagtcgtcga.....taatacgcgttcgaatrga 1140

Scoring table: OLIGO NUC

Gapop 60.0 , Gapext 60.0

Searched: 569978 seqs, 220691566 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Issued Patents NA:*

1: /cgn2_6/prodata/2/ina/5A_COMB.seq:*\n2: /cgn2_6/prodata/2/ina/5B_COMB.seq:*\n3: /cgn2_6/prodata/2/ina/6A_COMB.seq:*\n4: /cgn2_6/prodata/2/ina/6B_COMB.seq:*\n5: /cgn2_6/prodata/2/ina/6CTUS_COMB.seq:*\n6: /cgn2_6/prodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	19	1.7	4580	2	US-08-674-351-1
2	18	1.6	4975	2	US-08-249-687C-1
3	18	1.6	4989	2	US-08-666-392A-3
4	18	1.6	4989	2	US-08-625-819-1
5	18	1.6	4989	3	US-08-755-558-4
6	18	1.6	4989	3	US-08-746-559A-1
7	18	1.6	4989	3	US-08-880-313A-9
8	18	1.6	4989	3	US-09-199-926-3
9	18	1.6	4989	4	US-08-864-641B-17
10	18	1.6	4989	4	US-09-389-855A-9
11	18	1.6	4989	4	US-09-668-822-9
12	18	1.6	4993	3	US-08-746-559A-3
13	18	1.6	4403765	3	US-09-103-840A-2
14	18	1.6	4411529	3	US-09-103-840A-1
15	17	1.5	543	4	US-09-252-991A-12481
16	17	1.5	705	4	US-09-252-991A-12481
17	17	1.5	1653	3	US-09-230-944A-19
18	17	1.5	1653	3	US-08-881-784-8
19	17	1.5	1653	3	US-09-292-768-3
20	17	1.5	1653	3	US-09-292-768-67
21	17	1.5	1653	3	US-09-292-768-69
22	17	1.5	1687	1	US-08-279-700-17
23	17	1.5	1687	1	US-08-279-700-19
24	17	1.5	1762	3	US-08-881-784-5
25	17	1.5	1762	3	US-09-292-768-1
26	17	1.5	1762	3	US-09-292-768-63
27	17	1.5	1762	3	US-09-292-768-65

28	17	1.5	1762	3	US-09-172-339-5	Sequence 5, Appli
29	17	1.5	2352	4	US-08-997-251-3	Sequence 3, Appli
30	17	1.5	48908	4	US-09-453-702B-137	Sequence 137, App
31	17	1.5	4403765	3	US-09-103-840A-2	Sequence 2, Appli
32	17	1.5	4411529	3	US-09-103-840A-1	Sequence 1, Appli
33	16	1.4	139	4	US-09-141-027-11	Sequence 11, Appli
34	16	1.4	139	4	US-09-617-804-11	Sequence 11, Appli
35	16	1.4	219	2	US-08-414-657D-11	Sequence 11, Appli
36	16	1.4	219	2	US-08-414-657D-12	Sequence 11, Appli
37	16	1.4	276	4	US-09-673-809-23	Sequence 23, Appli
38	16	1.4	276	4	US-09-673-809-73	Sequence 73, Appli
39	16	1.4	340	2	US-08-997-362-137	Sequence 137, App
40	16	1.4	340	2	US-08-997-362-137	Sequence 137, App
41	16	1.4	340	3	US-09-095-855-137	Sequence 137, App
42	16	1.4	340	4	US-09-324-542-137	Sequence 137, App
43	16	1.4	340	4	US-09-205-426-137	Sequence 137, App
44	16	1.4	468	4	US-09-370-838-133	Sequence 133, App
45	16	1.4	504	4	US-09-134-001C-2229	Sequence 2229, Ap

ALIGNMENTS

RESULT 1
US-08-674-351-1/C
Sequence 1, Application US/08674351
Patent No. 5831013
GENERAL INFORMATION:
APPLICANT: Bruem, Jeremy A.
TITLE OF INVENTION: CAPSID POLYPEPTIDES AND USE TO INHIBIT
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSER: Nixon, Hargrave, Devans & Doyle LLP
STREET: Clinton Square, P.O. Box 1051
CITY: Rochester
STATE: New York
COUNTRY: U.S.A.
ZIP: 14603
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: IBM PC compatible
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/674,351
FILING DATE:
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Timian, Susan J.
REGISTRATION NUMBER: 34,103
REFERENCE/DOCKET NUMBER: 19226/740
TELECOMMUNICATION INFORMATION:
TELEPHONE: (716) 263-1636
TELEFAX: (716) 263-1600
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 4580 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULAR TYPE: CDNA
US-08-674-351-1
Query Match 1.7% Score 19; DB 2; Length 4580;
Best Local Similarity 100.0%; Pred. No. 4.2;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 541 GTGAGTGAACATCTCTCG 559
|||||
DB 3310 GTGAGTGAACATCTCTCG 3292

RESULT 2
US-08-249-687C-1/c
; Sequence 1, Application US/08249687C
; Patent No. 5942412
; GENERAL INFORMATION:
; APPLICANT: PRAGER, DIANE
; APPLICANT: MELMED, SHLOMO
; TITLE OF INVENTION: POLYNUCLEIC ACID ENCODING
; TITLE OF INVENTION: VARIANT INSULIN-LIKE GROWTH FACTOR I RECEPTOR BETA
; TITLE OF INVENTION: SUBUNIT & RECEPTOR
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pretty, Schroeder & Poplawski
; STREET: 444 South Flower St., 19th Floor
; CITY: Los Angeles
; STATE: CA
; COUNTRY: USA
; ZIP: 90071
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/249,687C
; FILING DATE: 26-MAY-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/044,540
; FILING DATE: 06-APR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Basile, Lena
; REGISTRATION NUMBER: P-44,026
; REFERENCE/DOCKET NUMBER: P07 32249
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 213-622-7700
; TELEFAX: 213-489-4210
; TELEX:
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4975 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: both
; MOLECULE TYPE: cDNA
US-08-249-687C-1

Query Match 1.6%; Score 18; DB 2; Length 4975;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1039 CTCGTCGAAGCTGGCGCG 1056
DB 4057 CTCGTCGAAGCTGGCGCG 4040

RESULT 3
US-08-666-392A-3/c
; Sequence 3, Application US/08666392A
; Patent No. 5928040
; GENERAL INFORMATION:
; APPLICANT: (countries other than U.S.): Royal Children's Hospital
; APPLICANT: Research Foundation
; APPLICANT: (U.S. only): George A. WERTHER and
; APPLICANT: Christopher J. WRIGHT
; TITLE OF INVENTION: A METHOD FOR THE PROPHYLAXIS AND/OR
; TITLE OF INVENTION: TREATMENT OF PROLIFERATIVE AND/OR
; TITLE OF INVENTION: INFLAMMATORY SKIN DISORDERS
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Brumbaugh, Graves, Donohue & Raymond
; STREET: 30 Rockefeller Plaza

CITY: New York
STATE: NY
COUNTRY: U.S.A.
ZIP: 10112-0228
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/666,392A
FILING DATE: 20-AUG-1996
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/AU95/00410
FILING DATE: 06-JUL-1995
ATTORNEY/AGENT INFORMATION:
NAME: Macleod, Janet M.
REGISTRATION NUMBER: 35,263
REFERENCE/DOCKET NUMBER: A30626-PCT-USA
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-408-2500
TELEFAX: 212-765-2519
TELEX:
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 4989 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE:
ORIGINAL SOURCE:
PUBLICATION INFORMATION:
AUTHORS: Ulrich, A., et al.
TITLE: Insulin-like growth factor I receptor...
JOURNAL: EMBO J.
VOLUME: 5
ISSUE: 1986
PAGES: 2503-2512
DATE:
US-08-666-392A-3

Query Match 1.6%; Score 18; DB 2; Length 4989;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1039 CTCGTCGAAGCTGGCGCG 1056
DB 4071 CTCGTCGAAGCTGGCGCG 4054

RESULT 4
US-08-625-819-1/c
; Sequence 1, Application US/08625819
; Patent No. 5958872
; GENERAL INFORMATION:
; APPLICANT: O'CONNOR, Rosemary; and
; APPLICANT: BASERGA, Renato L.
; TITLE OF INVENTION: ACTIVE SURVIVAL DOMAINS OF IGF-IR
; TITLE OF INVENTION: AND METHODS OF USE
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: HALE and DORR LLP
; STREET: 1455 Pennsylvania Avenue, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk

```

? COMPUTER: IBM PC compatible
? OPERATING SYSTEM: PC-DOS/MS-DOS
? SOFTWARE: Patent Release #1.0, Version #1.30
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08/625,819
? FILING DATE: 01-APR-1996
? CLASSIFICATION: 435
? ATTORNEY/AGENT INFORMATION:
? NAME: WIXON, Henry N.
? REGISTRATION NUMBER: 32,073
? REFERENCE/DOCKET NUMBER: 104322.162
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: (202) 942-8459
? TELEFAX: (202) 942-8484
? INFORMATION FOR SEQ ID NO: 1:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 4989 base pairs
? TYPE: nucleic acid
? STRANDEDNESS: single
? TOPOLOGY: linear
? FEATURE:
? NAME/KEY: CDS
? LOCATION: 46..4149
? FEATURE:
? NAME/KEY: mat_peptide
? LOCATION: 136..4149
? US-08-625-819-1

Query Match 1.6%; Score 18; DB 2; Length 4989;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1039 CTCGTGAAGCTGGCGCG 1056
Db 4071 CTCGTGAAGCTGGCGCG 4054

RESULT 5
US-08-755-558-4/C
? Sequence 4, Application US/08755558
? Patent No. 6071891
? GENERAL INFORMATION:
? APPLICANT: Low, Walter
? APPLICANT: Wallentriedman, Margaret
? APPLICANT: Chiang, Lan
? TITLE OF INVENTION: INSULIN-LIKE GROWTH FACTOR 1 RECEPTOR (IGF-1R) ANTISEN
? NUMBER OF SEQUENCES: 4
? CORRESPONDENCE ADDRESS:
? ADDRESSEE: Merchant, Gould, Smith, Edell, Welter & Schmidt
? STREET: 3100 No. 6071891west Center, 90 S. 7th Street
? CITY: Minneapolis
? STATE: MN
? COUNTRY: U.S.A.
? ZIP: 55402
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Diskette
? COMPUTER: IBM Compatible
? OPERATING SYSTEM: DOS
? SOFTWARE: FastSeq Version 1.5
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08/755,558
? FILING DATE: 20-NOV-1996
? CLASSIFICATION: 424
? PRIOR APPLICATION NUMBER:
? FILING DATE:
? ATTORNEY/AGENT INFORMATION:
? NAME: Kettlerberger, Denise M
? REGISTRATION NUMBER: 33,924
? REFERENCE/DOCKET NUMBER: 600.337US01
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: 612/332-5300
? TELEFAX: 612/332-9081
```

```

? TELEX:
? INFORMATION FOR SEQ ID NO: 4:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 4989 base pairs
? TYPE: nucleic acid
? STRANDEDNESS: single
? TOPOLOGY: linear
? MOLECULE TYPE: Genomic DNA
? HYPOTHETICAL: NO
? ANTI-SENSE: NO
? FRAGMENT TYPE:
? ORIGINAL SOURCE:
? FEATURE:
? NAME/KEY: Coding Sequence
? LOCATION: 46...4149
? OTHER INFORMATION:
? US-08-755-558-4

Query Match 1.6%; Score 18; DB 3; Length 4989;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```

QY 1039 CTCGTGAAGCTGGCGCG 1056
Db 4071 CTCGTGAAGCTGGCGCG 4054

RESULT 6
US-08-746-559A-1/C
? Sequence 1, Application US/08746559A
? Patent No. 6084085
? GENERAL INFORMATION:
? APPLICANT: Renato Baserga
? APPLICANT: Mariana Resnicoff
? APPLICANT: Consuelo D'Ambrosio
? APPLICANT: Andre Ferber
? TITLE OF INVENTION: Method of Inducing Resistance to Tumor Growth
? NUMBER OF SEQUENCES: 7
? CORRESPONDENCE ADDRESS:
? ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz & No. 6084085r1s LLP
? STREET: One Liberty Place - 46th Floor
? CITY: Philadelphia
? STATE: PA
? COUNTRY: USA
? ZIP: 19103
? COMPUTER READABLE FORM:
? MEDIUM TYPE: DISKETTE, 3.5 INCH, 1.44 Mb STORAGE
? COMPUTER: IBM PS/2
? OPERATING SYSTEM: PC-DOS
? SOFTWARE: WORDPERFECT 6.1
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08/746,559A
? FILING DATE: 13-NOV-1996
? CLASSIFICATION: 435
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: 60/006,699
? FILING DATE: 14-NOV-1995
? ATTORNEY/AGENT INFORMATION:
? NAME: Paul K. Legaard
? REGISTRATION NUMBER: 38,534
? REFERENCE/DOCKET NUMBER: TUT-2063
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: (215) 568-3100
? TELEFAX: (215) 568-3439
? INFORMATION FOR SEQ ID NO: 1:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 4989 base pairs
? TYPE: nucleic acid
? STRANDEDNESS: double
? TOPOLOGY: linear
? US-08-746-559A-1

Query Match 1.6%; Score 18; DB 3; Length 4989;
```

Best Local Similarity 100.0%; Pred. No. 14;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1039 CTCGTGAAGCTGGCGG 1056
Db 4071 CTCGTGAAGCTGGCGG 4054

```
RESULT 7
US-08-880-313A-9/c
; Sequence 9, Application US/08880313A
; Patent No. 6274562
; GENERAL INFORMATION:
; APPLICANT: Baseerga, Renato
; APPLICANT: Sell, Christian
; APPLICANT: Rubin, Raphael
; TITLE OF INVENTION: Method of Inhibiting the Proliferation and
; TITLE OF INVENTION: Causing the Differentiation of Cells with IGF-1 Receptor
; TITLE OF INVENTION: Antisense Oligonucleotides
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Woodcock Washburn Kurtz MacKiewicz & No. 6274562ris LLP
; STREET: One Liberty Place - 46th Floor
; CITY: Philadelphia
; STATE: PA
; COUNTRY: U.S.A.
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WordPerfect 8.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/880,313A
; FILING DATE: June 20, 1997
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/479,173
; FILING DATE: June 6, 1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Paul K. Legaard
; REGISTRATION NUMBER: 38,534
; REFERENCE/DOCKET NUMBER: TUD-2383
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-568-3100
; TELEFAX: 215-568-3439
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4989 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; US-08-880-313A-9

Query Match 1.6%; Score 18; DB 3; Length 4989;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1039 CTCGTGAAGCTGGCGG 1056
Db 4071 CTCGTGAAGCTGGCGG 4054

RESULT 8
US-09-199-926-3/c
; Sequence 3, Application US/09199926
; Patent No. 6284741
; GENERAL INFORMATION:
; APPLICANT: (countries other than U.S.): Royal Children's Hospital
; APPLICANT: Research Foundation
; APPLICANT: (U.S. only): George A. WERTHER and
; APPLICANT: Christopher J. WRIGHT
; TITLE OF INVENTION: A METHOD FOR THE PROPHYLAXIS AND/OR
```

```
; TITLE OF INVENTION: TREATMENT OF PROLIFERATIVE AND/OR
; TITLE OF INVENTION: INFLAMMATORY SKIN DISORDERS
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Brumbaugh, Graves, Donohue & Raymond
; STREET: 30 Rockefeller Plaza
; CITY: New York
; STATE: NY
; COUNTRY: U.S.A.
; ZIP: 10112-0228
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/199,926
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/666,392
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: MacLeod, Janet M.
; REGISTRATION NUMBER: 35,263
; REFERENCE/DOCKET NUMBER: A30626-PCT-USA
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-408-2500
; TELEFAX: 212-765-2519
; TELEX:
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4989 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE:
; ORIGINAL SOURCE:
; PUBLICATION INFORMATION:
; AUTHORS: Ulrich, A., et al.
; TITLE: Insulin-like growth factor I receptor...
; JOURNAL: EMBO J.
; VOLUME: 5
; ISSUE: 1986
; PAGES: 2503-2512
; DATE:
; US-09-199-926-3

Query Match 1.6%; Score 18; DB 3; Length 4989;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1039 CTCGTGAAGCTGGCGG 1056
Db 4071 CTCGTGAAGCTGGCGG 4054

RESULT 9
US-08-864-641B-17/c
; Sequence 17, Application US/08864641B
; Patent No. 6312684
; GENERAL INFORMATION:
; APPLICANT: Baseerga, Renato
; APPLICANT: Abraham, David
; APPLICANT: Resnickoff, Mariana
; TITLE OF INVENTION: Method of Inducing Resistance To Tumor Growth
; FILE REFERENCE: TUD2137
; CURRENT APPLICATION NUMBER: US/08/864,641B
; CURRENT FILING DATE: 1997-05-29
; PRIOR APPLICATION NUMBER: 08/340,732
```

PRIOR FILING DATE: 1994-11-16
NUMBER OF SEQ ID NOS: 18
SOFTWARE: PatentIn version 3.0
SEQ ID NO 17
LENGTH: 4989
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
NAME/KEY: misc feature
OTHER INFORMATION: No. 6312684el Sequence
NAME/KEY: CDS
LOCATION: (46)..(4146)
US-08-864-641B-17

Query Match 1.6%; Score 18; DB 4; Length 4989;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1039 CTCGTCGAAGCTGGCGCG 1056
DB 4071 CTCGTCGAAGCTGGCGCG 4054

RESULT 10
US-09-389-855A-9/c
Sequence 9, Application US/09389855A
Patent No. 6331526
GENERAL INFORMATION:

APPLICANT: Baseerga, Renato
APPLICANT: Sell, Christian
APPLICANT: Rubin, Raphael
TITLE OF INVENTION: Method Of Inhibiting The Proliferation And Causing The
TITLE OF INVENTION: Differentiation Of Cells With IGF-1 Receptor Antisense
TITLE OF INVENTION: Oligonucleotides
FILE REFERENCE: TJU-2388
CURRENT APPLICATION NUMBER: US/09/389,855A
CURRENT FILING DATE: 1999-09-02
PRIOR APPLICATION NUMBER: US 08/880,313
PRIOR FILING DATE: 1997-06-20
PRIOR APPLICATION NUMBER: US 08/479,173
PRIOR FILING DATE: 1995-06-06
PRIOR APPLICATION NUMBER: US 08/158,761
PRIOR FILING DATE: 1993-11-30
PRIOR APPLICATION NUMBER: US 08/037,257
PRIOR FILING DATE: 1993-03-26
NUMBER OF SEQ ID NOS: 9
SEQ ID NO 9
LENGTH: 4989
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:

OTHER INFORMATION: Other Information: human IGF-1R sequence
US-09-389-855A-9

Query Match 1.6%; Score 18; DB 4; Length 4989;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1039 CTCGTCGAAGCTGGCGCG 1056
DB 4071 CTCGTCGAAGCTGGCGCG 4054

RESULT 11
US-09-668-822-9/c
Sequence 9, Application US/09668822
Patent No. 6340674
GENERAL INFORMATION:

APPLICANT: Baseerga, Renato
APPLICANT: Sell, Christian
APPLICANT: Rubin, Raphael
TITLE OF INVENTION: Method of Inhibiting the Proliferation and Causing the Differ
NUMBER OF SEQUENCES: 9

CORRESPONDENCE ADDRESS:
ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz & No. 6340674rls LLP
STREET: One Liberty Place - 46th Floor
CITY: Philadelphia
STATE: PA
COUNTRY: U.S.A.
ZIP: 19103

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Wordperfect 8.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/668,822
FILING DATE: Herewith
CLASSIFICATION:

PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/880,313
FILING DATE: June 20, 1997
ATTORNEY/AGENT INFORMATION:

NAME: Paul K. Legaard
REGISTRATION NUMBER: 38,534
REFERENCE/DOCKET NUMBER: TJU-2443
TELECOMMUNICATION INFORMATION:
TELEPHONE: 215-568-3100
TELEFAX: 215-568-3439
INFORMATION FOR SEQ ID NO: 9:

SEQUENCE CHARACTERISTICS:
LENGTH: 4989 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
US-09-668-822-9

Query Match 1.6%; Score 18; DB 4; Length 4989;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1039 CTCGTCGAAGCTGGCGCG 1056
DB 4071 CTCGTCGAAGCTGGCGCG 4054

RESULT 12
US-08-746-559A-3/c
Sequence 3, Application US/08746559A
Patent No. 6084085
GENERAL INFORMATION:

APPLICANT: Renato Baseerga
APPLICANT: Mariana Resnicoff
APPLICANT: Consuelo D'Ambrosio
APPLICANT: Andre Ferber
TITLE OF INVENTION: Method of Inducing Resistance to Tumor Growth
NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESS:
ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz & No. 6084085rls LLP
STREET: One Liberty Place - 46th Floor
CITY: Philadelphia
STATE: PA
COUNTRY: USA
ZIP: 19103

COMPUTER READABLE FORM:
MEDIUM TYPE: DISKETTE, 3.5 INCH, 1.44 MB STORAGE
COMPUTER: IBM PS/2
OPERATING SYSTEM: PC-DOS
SOFTWARE: WORDPERFECT 6.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/746,559A
FILING DATE: 13-NOV-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/006,699
FILING DATE: 14-NOV-1995

```
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Paul K. Legaard
/ REGISTRATION NUMBER: 38,534
/ REFERENCE/DOCKET NUMBER: TJU-2063
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (215) 568-3100
/ TELEFAX: (215) 568-3439
/ INFORMATION FOR SEQ ID NO: 3:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 4993 base pairs
/ TYPE: nucleic acid
/ STRANDEDNESS: double
/ TOPOLOGY: linear
/ US-08-746-559A-3

Query Match          1.6%; Score 18; DB 3; Length 4993;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1039 CTCGTGCAAGCTGGCGC 1056
        |||||
Db       4075 CTCGTGCAAGCTGGCGC 4058

RESULT 13
US-09-103-840A-2
/ Sequence 2, Application US/09103840A
/ Patent No. 6294328
/ GENERAL INFORMATION:
/ APPLICANT: FLEISCHMAN, Robert D.
/ APPLICANT: WHITE, Owen R.
/ APPLICANT: FRASER, Claire M.
/ APPLICANT: VENTER, John C.
/ TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
/ TITLE OF INVENTION: TUBERCULOSIS
/ FILE REFERENCE: 24366-20007.00
/ CURRENT APPLICATION NUMBER: US/09/103,840A
/ CURRENT FILING DATE: 1998-06-24
/ NUMBER OF SEQ ID NOS: 2
/ SOFTWARE: PatentIn Ver. 2.1
/ SEQ ID NO 2
/ LENGTH: 4403765
/ TYPE: DNA
/ ORGANISM: Mycobacterium tuberculosis
/ FEATURE:
/ OTHER INFORMATION: CDC 1551
/ OTHER INFORMATION: "n" bases at various positions throughout the sequence
/ OTHER INFORMATION: represent a, t, c or g
/ US-09-103-840A-2

Query Match          1.6%; Score 18; DB 3; Length 4403765;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      759 ACATGGCAACCGTGACCC 776
        |||||
Db       2656723 ACATGGCAACCGTGACCC 2656740

RESULT 14
US-09-103-840A-1
/ Sequence 1, Application US/09103840A
/ Patent No. 6294328
/ GENERAL INFORMATION:
/ APPLICANT: FLEISCHMAN, Robert D.
/ APPLICANT: WHITE, Owen R.
/ APPLICANT: FRASER, Claire M.
/ APPLICANT: VENTER, John C.
/ TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
/ TITLE OF INVENTION: TUBERCULOSIS
/ FILE REFERENCE: 24366-20007.00
/ CURRENT APPLICATION NUMBER: US/09/103,840A
/ CURRENT FILING DATE: 1998-06-24
```

```
/ NUMBER OF SEQ ID NOS: 2
/ SOFTWARE: PatentIn Ver. 2.1
/ SEQ ID NO 1
/ LENGTH: 4411529
/ TYPE: DNA
/ ORGANISM: Mycobacterium tuberculosis
/ OTHER INFORMATION: H37Rv
/ US-09-103-840A-1

Query Match          1.6%; Score 18; DB 3; Length 4411529;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      759 ACATGGCAACCGTGACCC 776
        |||||
Db       2659427 ACATGGCAACCGTGACCC 2659444

RESULT 15
US-09-252-991A-12481
/ Sequence 12481, Application US/09252991A
/ Patent No. 6551795
/ GENERAL INFORMATION:
/ APPLICANT: Marc J. Rubenfield et al.
/ TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
/ TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
/ FILE REFERENCE: 107196.136
/ CURRENT APPLICATION NUMBER: US/09/252,991A
/ CURRENT FILING DATE: 1999-02-18
/ PRIOR APPLICATION NUMBER: US 60/074,788
/ PRIOR FILING DATE: 1998-02-18
/ PRIOR APPLICATION NUMBER: US 60/094,190
/ PRIOR FILING DATE: 1998-07-27
/ NUMBER OF SEQ ID NOS: 33142
/ SEQ ID NO 12481
/ LENGTH: 543
/ TYPE: DNA
/ ORGANISM: Pseudomonas aeruginosa
/ US-09-252-991A-12481

Query Match          1.5%; Score 17; DB 4; Length 543;
Best Local Similarity 100.0%; Pred. No. 47;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      132 CGCCTGTGGCGGCGCGA 148
        |||||
Db       422 CGCCTGTGGCGGCGCGA 438
```

Search completed: November 8, 2003, 23:04:09
Job time : 87 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 8, 2003, 21:35:22 ; Search time 304 Seconds
(without alignments)
11966.092 Million cell updates/sec

Title: US-10-007-527a-1

Perfect score: 1140
1 atgaccagcgtaagtcgta.....taatatcggttcgcaagtga 1140

Scoring table: OLIGO NUC
Gapop 60.0 , Gapext 60.0

Searched: 2141354 seqs, 1595478879 residues

Word size : 0

Total number of hits satisfying chosen parameters: 4282708

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database : Published Applications NA.*

1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq.*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq.*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq.*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq.*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq.*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq.*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq.*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq.*
9: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq.*
10: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq.*
11: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq.*
12: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq.*
13: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq.*
14: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq.*
15: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq.*
16: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq.*
17: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	length	ID	Description
1	1140	100.0	1140	US-10-007-527a-1	Sequence 1, Appli
2	1140	100.0	1140	US-10-007-452-1	Sequence 1, Appli
3	1140	100.0	6334	US-10-007-527a-5	Sequence 5, Appli
4	1140	100.0	6334	US-10-007-452-5	Sequence 5, Appli
5	1140	100.0	9652	US-10-007-527a-7	Sequence 7, Appli
6	1140	100.0	9652	US-10-007-452-7	Sequence 7, Appli
7	1140	100.0	11241	US-10-007-527a-6	Sequence 6, Appli
8	1140	100.0	11241	US-10-007-452-6	Sequence 6, Appli
9	20	1.8	20	US-10-007-527a-19	Sequence 19, Appli
10	20	1.8	20	US-10-007-452-19	Sequence 19, Appli
11	20	1.8	454	US-10-027-632-94798	Sequence 94798, A
12	20	1.8	454	US-10-027-632-94798	Sequence 94798, A
13	20	1.8	2055	US-10-225-810-5	Sequence 5, Appli
14	20	1.8	2093	US-09-805-456-1	Sequence 1, Appli
15	20	1.8	46649	US-09-805-456-3	Sequence 3, Appli
16	20	1.8	77992	US-10-225-810-11	Sequence 11, Appli

17	20	1.8	250000	US-10-225-810-26	Sequence 26, Appli
18	19	1.7	416	US-09-783-590-7376	Sequence 7376, Ap
19	19	1.7	447	US-09-918-995-9021	Sequence 9021, Ap
20	19	1.7	476	US-09-878-574-4387	Sequence 4387, Ap
21	19	1.7	1333	US-10-106-698-1797	Sequence 1797, Ap
22	19	1.7	3810	US-09-880-107-3709	Sequence 3709, Ap
23	18	1.6	20	US-10-007-527a-20	Sequence 20, Appli
24	18	1.6	20	US-10-007-452-20	Sequence 20, Appli
25	18	1.6	282	US-09-878-574-7769	Sequence 7769, Ap
26	18	1.6	399	US-09-764-869-1382	Sequence 1382, Ap
27	18	1.6	399	US-10-091-504-1382	Sequence 1382, Ap
28	18	1.6	440	US-09-764-869-64	Sequence 64, Appli
29	18	1.6	440	US-10-091-504-64	Sequence 64, Appli
30	18	1.6	1095	US-09-815-242-6193	Sequence 6193, Ap
31	18	1.6	1214	US-10-378-393-4	Sequence 4, Appli
32	18	1.6	1995	US-10-378-393-6	Sequence 6, Appli
33	18	1.6	4989	US-09-870-759-119	Sequence 119, App
34	18	1.6	4989	US-09-791-196-3	Sequence 3, Appli
35	18	1.6	4989	US-10-007-926a-130	Sequence 130, App
36	18	1.6	4989	US-09-751-708a-119	Sequence 119, App
37	18	1.6	4989	US-10-177-293-225	Sequence 225, App
38	18	1.6	8858	US-10-378-393-1	Sequence 1, Appli
39	17	1.5	113	US-09-969-373-620	Sequence 620, App
40	17	1.5	253	US-09-923-876-1088	Sequence 1088, Ap
41	17	1.5	343	US-09-734-563-109	Sequence 109, App
42	17	1.5	384	US-09-878-574-1881	Sequence 1881, Ap
43	17	1.5	509	US-10-029-386-11554	Sequence 11554, A
44	17	1.5	550	US-09-991-936-1453	Sequence 1453, Ap
45	17	1.5	561	US-10-027-632-290199	Sequence 290199,

ALIGNMENTS

RESULT 1
US-10-007-527a-1
Sequence 1, Application US/10007527a
Publication No. US20030044807A1
GENERAL INFORMATION:
APPLICANT: Tomb, Jean-Francois
APPLICANT: Bramucci, Michael G.
APPLICANT: Cheng, Qiong
APPLICANT: Kostichka, Kristy N.
TITLE OF INVENTION: Rhodococcus Cloning and Expression Vectors
FILE REFERENCE: C11709 US NA
CURRENT APPLICATION NUMBER: US/10/007,527a
CURRENT FILING DATE: 2001-12-05
PRIOR APPLICATION NUMBER: 60/254,868
PRIOR FILING DATE: 2000-12-12
NUMBER OF SEQ ID NOS: 30
SOFTWARE: Microsoft Office 97
SEQ ID NO 1
LENGTH: 1140
TYPE: DNA
ORGANISM: Rhodococcus AN12
US-10-007-527a-1

Query Match 100.0%; Score 1140; DB 14; length 1140;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1140; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	ATGACCAGCGTAAGTCTGAACACTTTCCGGGAAAGACCGGCTCCCTCGTGTGCG 60
DB	1	ATGACCAGCGTAAGTCTGAACACTTTCCGGGAAAGACCGGCTCCCTCGTGTGCG 60
QY	61	TCGGATTAACGCGGCGATTCGGGACGAGCCGACCTTCAAGAAATCAACAGTCA 120
DB	61	TCGGATTAACGCGGCGATTCGGGACGAGCCGACCTTCAAGAAATCAACAGTCA 120
QY	121	GAACATTTAAAGCTGTGCGCGCGCGATTTCTGGCGTGAACGATGTGACATTTGCAAC 180
DB	121	GAACATTTAAAGCTGTGCGCGCGCGATTTCTGGCGTGAACGATGTGACATTTGCAAC 180

RESULT 2
US-10-007-452-1
; Sequence 1, Application US/10007452
; Publication No. US20030093701A1
; GENERAL INFORMATION:
; APPLICANT: Tomb, Jean-Francois
; APPLICANT: Bramucci, Michael G.
; APPLICANT: Cheng, Qiong

Query Match	100.0%;	Score 1140;	DB 14;	Length 1140;
Best Local Similarity	100.0%;	Pred. No. 0;		
Matches 1140; Conservative	0;	Mismatches	0;	Indels 0; Gaps 0;

OY	1	ATGACGAGCGTAAGTGTGTGAAACACCTTTCGGGCAAAAGACGGGCTCCGCTCTGTGTGCG	60
Db	1	ATGACGAGCGTAAGTGTGTGAAACACCTTTCGGGCAAAAGACGGGCTCCGCTCTGTGTGCG	60
OY	61	TCCGATAAGCGCGGCATCCGGACGGAACGTGGACCCAAATTTCAACAAATCAACAAGTCGA	120
Db	61	TCCGATAAGCGCGGCATCCGGACGGAACGTGGACCCAAATTTCAACAAATCAACAAGTCGA	120
OY	121	GAAACATTAAACGCTGTGGCCGGCCGATTTCTGGCGTGAACGGGTGTGACCAATTGTGCAC	180
Db	121	GAAACATTAAACGCTGTGGCCGGCCGATTTCTGGCGTGAACGGGTGTGACCAATTGTGCAC	180
OY	181	GGTCCGAAAGGTTTGTGATTCGAGAGGCGCTTGTTCTCGCGAAGAGGCTGTGATCTGGCCC	240
Db	181	GGTCCGAAAGGTTTGTGATTCGAGAGGCGCTTGTTCTCGCGAAGAGGCTGTGATCTGGCCC	240
OY	241	TGCTGTGGCGGAAAGTCGGTGTGCAATCGTGTGACAGAAATTTCTCAAGTTGTGTGCAT	300
Db	241	TGCTGTGGCGGAAAGTCGGTGTGCAATCGTGTGACAGAAATTTCTCAAGTTGTGTGCAT	300
OY	301	CAACTCGGGACTGATCTGTGGATGGTGCATGACATGACCATGGCCATACAGCTGTGAG	360
Db	301	CAACTCGGGACTGATCTGTGGATGGTGCATGACATGACCATGGCCATACAGCTGTGAG	360
OY	361	CGGCTCCACGACCTATGTGACTGTTCGGGACCGCTGGAAAGCGTGGACCAACGAGTGTG	420
Db	361	CGGCTCCACGACCTATGTGACTGTTCGGGACCGCTGGAAAGCGTGGACCAACGAGTGTG	420
OY	421	CGTTGACGCTACGGAACGTGAATATATACGGCTGGACAGCATACGAGTGTGGAATTC	480
Db	421	CGTTGACGCTACGGAACGTGAATATATACGGCTGGACAGCATACGAGTGTGGAATTC	480
OY	481	ACTCACGGAAGAAACGCTGGCACGTCCACGTTCAACGCGCTACTCATGTTCAAGTGTGAC	540
Db	481	ACTCACGGAAGAAACGCTGGCACGTCCACGTTCAACGCGCTACTCATGTTCAAGTGTGAC	540
OY	541	GTGAGTGAACAATCTTCGAATCCTTCTTCGGAATGCGATGCGATTCGCGTTCACAA	600
Db	541	GTGAGTGAACAATCTTCGAATCCTTCTTCGGAATGCGATGCGATTCGCGTTCACAA	600
OY	601	CTCGATATCTCGGAATTTGCTGGCGCACTAGAGTAATGGGGGGTCTCGATGTAGAAAG	660
Db	601	CTCGATATCTCGGAATTTGCTGGCGCACTAGAGTAATGGGGGGTCTCGATGTAGAAAG	660
OY	661	ATCGGCGGTGAAGCTGATCAAGTTCTCGCTCGTATCTGACGAAATTTGCATCTGCCTT	720
Db	661	ATCGGCGGTGAAGCTGATCAAGTTCTCGCTCGTATCTGACGAAATTTGCATCTGCCTT	720
OY	721	GGTATGAGAGTTGTGATGTGGCGACGGAAGAAAGTGTGTGACATGGCAACCGTACCTTGG	780
Db	721	GGTATGAGAGTTGTGATGTGGCGACGGAAGAAAGTGTGTGACATGGCAACCGTACCTTGG	780
OY	781	GAAATCGCTGTGATGACGTGGCGGGGATCCACAAGGTTGAAACTGTGGGGAATAATT	840
Db	781	GAAATCGCTGTGATGACGTGGCGGGGATCCACAAGGTTGAAACTGTGGGGAATAATT	840

Db	781	GAATTCGCTGTGATGACAGTGGGAGGATATCCAAAGAGTTTGAACTGTGGCGAGAAATT	840
Qy	841	GAGTTTGGTTCGATATGGGACGTCCGGGCAATGCGCGTGTCCGTGTGATGTCCGGTCCCGAGCT	900
Db	841	GAGTTTGGTTCGATATGGGACGTCCGGGCAATGCGCGTGTCCGTGTGATGTCCGGTCCCGAGCT	900
Qy	901	GGTCTTTGGGGCAGAACTAACAGATGCTCAGATCGTTGAGCAGAAAGATCTGCCCGGCTC	960
Db	901	GGTCTTTGGGGCAGAACTAACAGATGCTCAGATCGTTGAGCAGAAAGATCTGCCCGGCTC	960
Qy	961	ATGATTGCGATCATTTCCGGCGCGATCGTGGATATATGATTCGGACCTTGTGGCCCTTAAGTC	1020
Db	961	ATGATTGCGATCATTTCCGGCGCGATCGTGGATATATGATTCGGACCTTGTGGCCCTTAAGTC	1020
Qy	1021	TTCCGCGAGATCCTCGACATCGTGGAAAGCTGGCGACTTGGGAAAAATCTTTCGTGATCAC	1080
Db	1021	TTCCGCGAGATCCTCGACATCGTGGAAAGCTGGCGACTTGGGAAAAATCTTTCGTGATCAC	1080
Qy	1081	TTTGCAATTATCGATTGGCCCGCAGCGGAGATGTGGCGCCCGGAGAAATATTCGTTCCGAAATGA	1140
Db	1081	TTTGCAATTATCGATTGGCCCGCAGCGGAGATGTGGCGCCCGGAGAAATATTCGTTCCGAAATGA	1140

```

RESULT 3
US-10-007-527A-5/c
; Sequence 5, Application US/10007527A
; Publication No. US20030044807A1
; GENERAL INFORMATION:
; APPLICANT: Tomb, Jean-Francois
; APPLICANT: Bramucci, Michael G.
; APPLICANT: Cheng, Qiong
; APPLICANT: Kostichka, Kristy N.
; TITLE OF INVENTION: Rhodococcus Cloning and Expression Vectors
; FILE REFERENCE: C11709 US NA
; CURRENT APPLICATION NUMBER: US/10/007,527A
; CURRENT FILING DATE: 2001-12-05
; PRIOR APPLICATION NUMBER: 60/254,868
; PRIOR FILING DATE: 2000-12-12
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 5
; LENGTH: 6334
; TYPE: DNA
; ORGANISM: Rhodococcus AN12
US-10-007-527A-5

```

Query Match	100.0%;	Score 1140;	DB 14;	Length 6334;
Best Local Similarity	100.0%;	Pred. No. 0;		
Matches 1140;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;

OY	1	ATGACCAAGGCTAAGTGCCTGAA	CACCTTTCCGGCAAAAGACGGGCTCCGGTCTCGTGTGCG	60
Db	3051	ATGACCAAGGCTAAGTGCCTGAA	CACCTTTCCGGCAAAAGACGGGCTCCGGTCTCGTGTGCG	2992
OY	61	TCGGATAAGCGCGCATCCGSCACGAA	CTGCGACCTGAACCTTGAACAAATACACACGTCA	120
Db	2991	TCGGATAAGCGCGCATCCGSCACGAA	CTGCGACCTGAACCTTGAACAAATACACACGTCA	2932
OY	121	GAAACATTTAAACGCGCTGTGAGCCGGCGCGAATTTCTGGCGTAAACGAGTGTGACATTTGTCAAC		180
Db	2931	GAAACATTTAAACGCGCTGTGAGCCGGCGCGAATTTCTGGCGTAAACGAGTGTGACATTTGTCAAC		2872
OY	181	GGTCCGAAAGAGTTCTGGATTCGAGAGGCGCTTCGTTCTCTCGGAAAGGCGCTGGATCTGCCCC		240
Db	2871	GGTCCGAAAGAGTTCTGGATTCGAGAGGCGCTTCGTTCTCTCGGAAAGGCGCTGGATCTGCCCC		2811
OY	241	TGCTGTGCGGGGAAAAGTGGGTGCA	CATCGTGCAGACGAATTTCTCAAGTTGTTCCTCAT	300
Db	2811	TGCTGTGCGGGGAAAAGTGGGTGCA	CATCGTGCAGACGAATTTCTCAAGTTGTTCCTCAT	275
OY	301	CAACTCGGAGCTGGATCTTGTTCGATG	GTGAGAGATGACCATGCGGCATACACCTGGTGTAG	360
Db	2751	CAACTCGGAGCTGGATCTTGTTCGATG	GTGAGAGATGACCATGCGGCATACACCTGGTGTAG	269

QY	361	CGGCTCCAGACACCTATGGA	CTGGACCTTTGGGACAG	CCCTGAAAGAGCTGGGAC	CCAAAGGTCGT	420
Db	2691	CGGCTCCAGACACCTATGGA	CTGGACCTTTGGGACAG	CCCTGAAAGAGCTGGGAC	CCAAAGGTCGT	2632
QY	421	CGTTGGCGTACGGAA	CGTGAATGTACAGGCTG	CGACGGATACGTCGCGCTGTGAAATC		480
Db	2631	CGTTGGCGGTACGGAA	CGTGAATGTACAGGCTG	CGACGGATACGTCGCGCTGTGAAATC		2572
QY	481	ACTCACGGAAAAA	CGGCTGGCACGTC	CCAGTTTACAGCGCTACTCATGTTCAAGTGTGAC	540	
Db	2571	ACTCACGGAAAAA	CGGCTGGCACGTC	CCAGTTTACAGCGCTACTCATGTTCAAGTGTGAC	2512	
QY	541	GTGAGTGAAGAA	CAATCCCTCGAATCCCTTCGAGTACGAA	AGTTTCGATCCGCTGGACCTTCCAA	600	
Db	2511	GTGAGTGAAGAA	CAATCCCTCGAATCCCTTCGAGTACGAA	AGTTTCGATCCGCTGGACCTTCCAA	2455	
QY	601	CTCGATATCTCGGAGTAT	TGCTGCGCCACACTAGTAAT	CGGCTGGTCTCGATGTACGAAG	660	
Db	2451	CTCGATATCTCGGAGTAT	TGCTGCGCCACACTAGTAAT	CGGCTGGTCTCGATGTACGAAG	2392	
QY	661	ATCGGCGGTGAAGCTGAT	CAAATTCTGCTGCGTATCTGACGAA	AATTGCAATCTGGCGTT	720	
Db	2391	ATCGGCGGTGAAGCTGAT	CAAATTCTGCTGCGTATCTGACGAA	AATTGCAATCTGGCGTT	2333	
QY	721	GGTATGAGAGGTTGAT	GAGTGGGACAGGAAAAAGTGTGCA	ATGGCAACGAGGACCCCGG	780	
Db	2331	GGTATGAGAGGTTGAT	GAGTGGGACAGGAAAAAGTGTGCA	ATGGCAACGAGGACCCCGG	2272	
QY	781	GAAATCGCTGTGATG	CAGTGGGCGGGGATCCAAAGGCTGGA	ACTGTGCGAGAAATTT	840	
Db	2271	GAAATCGCTGTGATG	CAGTGGGCGGGGATCCAAAGGCTGGA	ACTGTGCGAGAAATTT	2211	
QY	841	GAGTTTGTTGCATGGAG	CGTGGGCAATCGCGTGTCCCGTGAT	TGCGTGGCCGAGCT	900	
Db	2211	GAGTTTGTTGCATGGAG	CGTGGGCAATCGCGTGTCCCGTGAT	TGCGTGGCCGAGCT	2152	
QY	901	GGTCTTGGGGGAGAA	CTPACAGATGCTCAGATCGTTGACAGGAA	AGATCTGSCCCCGGTC	960	
Db	2151	GGTCTTGGGGGAGAA	CTPACAGATGCTCAGATCGTTGACAGGAA	AGATCTGSCCCCGGTC	2092	
QY	961	ATGGTTGGATCATCT	CGGCGCGCATCTGATGATGATTCGGA	ACTTGTGCGCCTTACGTC	1020	
Db	2091	ATGGTTGGATCATCT	CGGCGCGCATCTGATGATGATTCGGA	ACTTGTGCGCCTTACGTC	2033	
QY	1021	TTGCGCAGGATCTT	CGAATCTGTCGAGCTGGCGACTTGGGAAA	AATCTTGTGATCAC	1080	
Db	2031	TTGCGCAGGATCTT	CGAATCTGTCGAGCTGGCGACTTGGGAAA	AATCTTGTGATCAC	1972	
QY	1081	TTTGATTTATGATTT	GCCCGGACGGAGTGTGGG	CCCCCGGATATATTCGGTTGCAAGTGA	1144	
Db	1971	TTTGATTTATGATTT	GCCCGGACGGAGTGTGGG	CCCCCGGATATATTCGGTTGCAAGTGA	1912	

```

1  RESULT 4
2  US-10-007-452-5/C
3  ; Sequence 5, Application US/10007452
4  ; Publication No. US20030093701A1
5  ; GENERAL INFORMATION:
6  ; APPLICANT: Tomb, Jean-Francois
7  ; APPLICANT: Bramucci, Michael G.
8  ; APPLICANT: Cheng, Qiong
9  ; APPLICANT: Kostichka, Kristy N.
10 ; TITLE OF INVENTION: Rhodococcus Cloning and Expression Vectors
11 ; FILE REFERENCE: CH1709 US NA
12 ; CURRENT APPLICATION NUMBER: US/10/007,452
13 ; CURRENT FILING DATE: 2001-11-08
14 ; PRIOR APPLICATION NUMBER: 60/254,868
15 ; PRIOR FILING DATE: 2000-12-12
16 ; NUMBER OF SEQ ID NOS: 30
17 ; SOFTWARE: Microsoft Office 97
18 ; SEQ ID NO 5
19 ; LENGTH: 6334

```

TYPE: DNA
ORGANISM: Rhodococcus AN12
US-10-007-452-5

Query Match 100.0%; Score 1140; DB 14; Length 6334;

Best Local Similarity 100.0%; Pred. No. 0; Mismatches 0; Indels 0; Gaps 0;

Matches 1140; Conservative 0; Indels 0; Gaps 0;

```
QY 1 ATGACGAGGTAAGTGTGTAACACCTTTCCGGCAAAAGACCGGCTCTCCGCTCTGTCG 60
DB 3051 ATGACGAGGTAAGTGTGTAACACCTTTCCGGCAAAAGACCGGCTCTCCGCTCTGTCG 2992

QY 61 TCCGATTAAGCGCGGATCCGCGACAACTGCGACCCAACTTCAACAAATCAACACGTC 120
DB 2991 TCCGATTAAGCGCGGATCCGCGACAACTGCGACCCAACTTCAACAAATCAACACGTC 2932

QY 121 GAAACATTTAAAGCTGTCGCGCGCGGCGGATTTCTGGCGGTAAGGTTGATGACCTTGTG 180
DB 2931 GAAACATTTAAAGCTGTCGCGCGCGGCGGATTTCTGGCGGTAAGGTTGATGACCTTGTG 2872

QY 181 GGTCCGAAAGTTCTGATTCGAGGCTTCTGTCCTCGGAAAGGCGCTGATCTGCCCC 240
DB 2871 GGTCCGAAAGTTCTGATTCGAGGCTTCTGTCCTCGGAAAGGCGCTGATCTGCCCC 2812

QY 241 TCGTGTGCGGAAAGTCTGTCGACATCTGTCGACAGCAAAATTTCTCAAGTTGTCTCAT 300
DB 2811 TCGTGTGCGGAAAGTCTGTCGACATCTGTCGACAGCAAAATTTCTCAAGTTGTCTCAT 2752

QY 301 CAACCTCGGAGCTGGAATCTGTGCGAATGTGACGATGACATGCGGCAATGACGCTGTCAG 360
DB 2751 CAACCTCGGAGCTGGAATCTGTGCGAATGTGACGATGACATGCGGCAATGACGCTGTCAG 2692

QY 361 CGGCTCCAGCACTATGACCTGATCTTTCGCGAGCTTGGAAAGCTGCGACCAACGCTGCT 420
DB 2691 CGGCTCCAGCACTATGACCTGATCTTTCGCGAGCTTGGAAAGCTGCGACCAACGCTGCT 2632

QY 421 CGTTGCGTAACGAAAGTGTGACATCTGTCGACGATGACGATGACGATGACGATGAC 480
DB 2631 CGTTGCGTAACGAAAGTGTGACATCTGTCGACGATGACGATGACGATGACGATGAC 2572

QY 481 ACTCAAGGAAAGAAACGCGTGGGACGTCGACGTCGACGCGCTACTCAATGTTCAAGTGTG 540
DB 2571 ACTCAAGGAAAGAAACGCGTGGGACGTCGACGTCGACGCGCTACTCAATGTTCAAGTGTG 2512

QY 541 GTGAGTGAAGAAACATCTCGAATCTTCTCGATGCGATGTTGATGCGTGTGCACTTCCAA 600
DB 2511 GTGAGTGAAGAAACATCTCGAATCTTCTCGATGCGATGTTGATGCGTGTGCACTTCCAA 2452

QY 601 CTGATCTCTGGAATTTGCTGCGCACTACGTAATTCGAGTGTCTCGATGTGCAAAAG 660
DB 2451 CTGATCTCTGGAATTTGCTGCGCACTACGTAATTCGAGTGTCTCGATGTGCAAAAG 2392

QY 661 ATGCGGCGTGAACCTGATCAAGTTCTGCGCGATCTGATGCAAGAAATTTGATGCGAGTT 720
DB 2391 ATGCGGCGTGAACCTGATCAAGTTCTGCGCGATCTGATGCAAGAAATTTGATGCGAGTT 2332

QY 721 GGTATGAGGTTGGTATGCGACGCGAAAGTGTGCAATGCGCAACCGTGTGCAACCTG 780
DB 2331 GGTATGAGGTTGGTATGCGACGCGAAAGTGTGCAATGCGCAACCGTGTGCAACCTG 2272

QY 781 GAAATGCTGTGATGATGAGTGGGCGGGAATTCACAGCGTTGGAATCTGTGGCGAATTT 840
DB 2271 GAAATGCTGTGATGATGAGTGGGCGGGAATTCACAGCGTTGGAATCTGTGGCGAATTT 2212

QY 841 GAGTTTGGTTCGATGAGACGTCGCGGCAATGCGGTGTCCTGATGTTGCGTGGCGAGCT 900
DB 2211 GAGTTTGGTTCGATGAGACGTCGCGGCAATGCGGTGTCCTGATGTTGCGTGGCGAGCT 2152

QY 901 GGTCTTGGGCGAAGTCAACAGATGCTCAAGTCTGTTAGCGAGAAAGATCTGCCCGGCTC 960
DB 2151 GGTCTTGGGCGAAGTCAACAGATGCTCAAGTCTGTTAGCGAGAAAGATCTGCCCGGCTC 2092

QY 961 ATGTTGCGATCAATCCGCGCGGATCGTGTGATGATGATTCGAACTTGTGCGCTTACGTC 1020
```

```
DB 2091 ATGTTTCCGATCAATCCGCGCGGATCGTGTGATGATGATTCGAACTTGTGCGCTTACGTC 2032
QY 1021 TTGCGGAGATCTCTGGAATCTGTGAAAGCTGGCGGCACTTGGGAAATCTTCTGTGATCAC 1080
DB 2031 TTGCGGAGATCTCTGGAATCTGTGAAAGCTGGCGGCACTTGGGAAATCTTCTGTGATCAC 1972
QY 1081 TTGATTTATGATTTGCGGCGAGCGGATGTCGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1140
DB 1971 TTGATTTATGATTTGCGGCGAGCGGATGTCGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1912
```

RESULT 5

US-10-007-527a-7/c

Sequence 7, Application US/1007527A

Publication No. US20030044807A1

GENERAL INFORMATION:

APPLICANT: Bammuel, Michael G.

APPLICANT: Cheng, Qiong

APPLICANT: Kostichka, Kristy N.

TITLE OF INVENTION: Rhodococcus Cloning and Expression Vectors

FILE REFERENCE: C11709 US NA

CURRENT FILING DATE: 2001-12-05

PRIOR APPLICATION NUMBER: 60/254,868

PRIOR FILING DATE: 2000-12-12

NUMBER OF SEQ ID NOS: 30

SOFTWARE: Microsoft Office 97

SEQ ID NO 7

LENGTH: 9652

TYPE: DNA

ORGANISM: Plasmid pRHR17

US-10-007-527a-7

Query Match 100.0%; Score 1140; DB 14; Length 9652;

Best Local Similarity 100.0%; Pred. No. 0; Mismatches 0; Indels 0; Gaps 0;

Matches 1140; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 ATGACGAGGTAAGTGTGTAACACCTTTCCGGCAAAAGACCGGCTCTCCGCTCTGTCG 60
DB 5052 ATGACGAGGTAAGTGTGTAACACCTTTCCGGCAAAAGACCGGCTCTCCGCTCTGTCG 4993

QY 61 TCCGATTAAGCGCGGATCCGCGACAACTGCGACCCAACTTCAACAAATCAACACGTC 120
DB 4992 TCCGATTAAGCGCGGATCCGCGACAACTGCGACCCAACTTCAACAAATCAACACGTC 4933

QY 121 GAAACATTTAAAGCTGTCGCGCGCGGCGGATTTCTGGCGGTAAGGTTGATGACCTTGT 180
DB 4932 GAAACATTTAAAGCTGTCGCGCGCGGCGGATTTCTGGCGGTAAGGTTGATGACCTTGT 4873

QY 181 GGTCCGAAAGTTCGATTCGAGGCTTCTGTCCTCGGAAAGGCGCTGATCTGCCCC 240
DB 4872 GGTCCGAAAGTTCGATTCGAGGCTTCTGTCCTCGGAAAGGCGCTGATCTGCCCC 4813

QY 241 TCGTGTGCGGAAAGTCTGTCGACATCTGTCGACAGCAATTTCTCAAGTTGTGTCAT 300
DB 4812 TCGTGTGCGGAAAGTCTGTCGACATCTGTCGACAGCAATTTCTCAAGTTGTGTCAT 4753

QY 301 CAACCTCGGAGCTGATCTGTCGATGTCGATGTCGATGTCGATGTCGATGTCGATGTCG 360
DB 4752 CAACCTCGGAGCTGATCTGTCGATGTCGATGTCGATGTCGATGTCGATGTCGATGTCG 4693

QY 361 CGGCTCCAGCACTATGACCTGATCTTTCGCGAGCTTGGAAAGCTGCGACCAACGCTGCT 420
DB 4692 CGGCTCCAGCACTATGACCTGATCTTTCGCGAGCTTGGAAAGCTGCGACCAACGCTGCT 4633

QY 421 CGTTGCGGTAACGAAAGTGTGACATCTGTCGACAGCAATGTCGACAGCAATGTCGAC 480
DB 4632 CGTTGCGGTAACGAAAGTGTGACATCTGTCGACAGCAATGTCGACAGCAATGTCGAC 4573

QY 481 ACTCAAGGAAAGAAACGCGTGGGACGTCAGTTCAGGCTACTCAATGTTCAAGTGTGAC 540
```


RESULT 7
US-10-007-527a-6/c
; Sequence 6, Application US/10007527A
; Publication No. US20030044807A1
; GENERAL INFORMATION:
; APPLICANT: Tomb, Jean-Francois
; APPLICANT: Bramucci, Michael G.
; APPLICANT: Cheng, Qiong
; APPLICANT: Kostichka, Kristy N.
; TITLE OF INVENTION: Rhodococcus Cloning and Expression Vectors
; REFERENCE: C11709 US NA
; CURRENT APPLICATION NUMBER: US/10/007,527A
; PRIOR FILING DATE: 2001-12-05
; PRIOR APPLICATION NUMBER: 60/254,868
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 6
; LENGTH: 11241
; TYPE: DNA
; ORGANISM: Plasmid pRHR17
US-10-007-527a-6

Query Match 100.0%; Score 1140; DB 14; Length 11241;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1140; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGACGAGGTAAGTCTGAACACCTTTCGCGAAGACCGGCTCCGTCCTGTCG 60
DB 6641 ATGACGAGGTAAGTCTGAACACCTTTCGCGAAGACCGGCTCCGTCCTGTCG 6582
QY 61 TCCGATTAAGCGCGCATCCGCGACGAACCTGCGAACCCAACTTCAACAAATCACACGTCA 120
DB 6581 TCCGATTAAGCGCGCATCCGCGACGAACCTGCGAACCCAACTTCAACAAATCACACGTCA 6522
QY 121 GAAACATTAAAGCCCTGTGGCCCGCGGATTTCTGGGGTGAACGCTGTGACATTTGTCAC 180
DB 6521 GAAACATTAAAGCCCTGTGGCCCGCGGATTTCTGGGGTGAACGCTGTGACATTTGTCAC 6462
QY 181 GGTCCGAAAGGTTCTGATTCGAGAGGCTTCTTCTCGCGAAGGGCTGATCTGCCCC 240
DB 6461 GGTCCGAAAGGTTCTGATTCGAGAGGCTTCTTCTCGCGAAGGGCTGATCTGCCCC 6402
QY 241 TGCTGTGCGGAAAAAGTCGTCGACATCGTCGACGAAATTTCTCAAGTTGTGTCAT 300
DB 6401 TGCTGTGCGGAAAAAGTCGTCGACATCGTCGACGAAATTTCTCAAGTTGTGTCAT 6342
QY 301 CAACTCGGACCTGATCTGTTGCGATGTGACATGACATGCGCCATTAAGCTGTGACG 360
DB 6341 CAACTCGGACCTGATCTGTTGCGATGTGACATGACATGCGCCATTAAGCTGTGACG 6282
QY 361 CGGCTCCAGACCTATGATCTGAGACTTTCGCGAGCTCGAAGAGCTGCAACCAAGCTGT 420
DB 6281 CGGCTCCAGACCTATGATCTGAGACTTTCGCGAGCTCGAAGAGCTGCAACCAAGCTGT 6222
QY 421 CGTTGGGTAAGGAACTGTAATGTAGAGCTGCGACGATACGTCGCGCTGTGAATC 480
DB 6221 CGTTGGGTAAGGAACTGTAATGTAGAGCTGCGACGATACGTCGCGCTGTGAATC 6162
QY 481 ACTCAGGAAAAAGCGCTGCGACGTCACGTTTACCGGCTACTCATGTTCACTGTGAC 540
DB 6161 ACTCAGGAAAAAGCGCTGCGACGTCACGTTTACCGGCTACTCATGTTCACTGTGAC 6102
QY 541 GTGAGTGAAGAACTCTCGAATCTTCTCGAGATGCGATGTTTCACTGCGTGAATTC 600
DB 6101 GTGAGTGAAGAACTCTCGAATCTTCTCGAGATGCGATGTTTCACTGCGTGAATTC 6042
QY 601 CTGATATCTCTGGGATTTGTCGCGCACTAGATTAATCGGGTGTCTCGATGTAAGAA 660
DB 6041 CTGATATCTCTGGGATTTGTCGCGCACTAGATTAATCGGGTGTCTCGATGTAAGAA 5982
QY 661 ATGCGGCGTGAAGCTGATCAAGTTCTGCGCGATCTGACGAAAAATTCGATCTGGGCT 720

DB 5981 ATGCGGCGTGAAGCTGATCAAGTTCTGCGCGATCTGACGAAAAATTCGATCTGGGCT 5922
QY 721 GGTATGAGGTTGTAGTGGCGACGAAAAAGTGTGACATGCGAACCTGACCCCTGG 780
DB 5921 GGTATGAGGTTGTAGTGGCGACGAAAAAGTGTGACATGCGAACCTGACCCCTGG 5862
QY 781 GAAATCGCTGTGATGCGATGCGGGGATCCAGAGGTTGAACCTGTGCGAGATTT 840
DB 5861 GAAATCGCTGTGATGCGATGCGGGGATCCAGAGGTTGAACCTGTGCGAGATTT 5802
QY 841 GAGTTGTGATGAGGACGTCGCGCAATCCGCTGTGATCCGCTGATTCGCGAGCT 900
DB 5801 GAGTTGTGATGAGGACGTCGCGCAATCCGCTGTGATCCGCTGATTCGCGAGCT 5742
QY 901 GGTCTTGGGCGAAGCTTAACAGATCTCAGATCTGTTAGAGAGAAATTCGCCGCT 960
DB 5741 GGTCTTGGGCGAAGCTTAACAGATCTCAGATCTGTTAGAGAGAAATTCGCCGCT 5682
QY 961 ATGTTTGGATCATTCGCGCGGATCGTGTGATGATGATTCGACTTGTGCGCTTACGTC 1020
DB 5681 ATGTTTGGATCATTCGCGCGGATCGTGTGATGATGATTCGACTTGTGCGCTTACGTC 5622
QY 1021 TTCGCGAGATCTCTGCACTCGTCGAGCTGCGCGCACTTGGGAAAAATCTGTGATCAC 1080
DB 5621 TTCGCGAGATCTCTGCACTCGTCGAGCTGCGCGCACTTGGGAAAAATCTGTGATCAC 5562
QY 1081 TTGCATTAATGATTCGCGCGAGGATGTCGCGCCCGGATTAATTCGTTTGCAGATGA 1140
DB 5561 TTGCATTAATGATTCGCGCGAGGATGTCGCGCCCGGATTAATTCGTTTGCAGATGA 5502

RESULT 8
US-10-007-452-6/c
; Sequence 6, Application US/10007452
; Publication No. US20030093701A1
; GENERAL INFORMATION:
; APPLICANT: Tomb, Jean-Francois
; APPLICANT: Bramucci, Michael G.
; APPLICANT: Cheng, Qiong
; APPLICANT: Kostichka, Kristy N.
; TITLE OF INVENTION: Rhodococcus Cloning and Expression Vectors
; REFERENCE: C11709 US NA
; CURRENT APPLICATION NUMBER: US/10/007,452
; PRIOR FILING DATE: 2001-11-08
; PRIOR APPLICATION NUMBER: 60/254,868
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 6
; LENGTH: 11241
; TYPE: DNA
; ORGANISM: Plasmid pRHR17
US-10-007-452-6

Query Match 100.0%; Score 1140; DB 14; Length 11241;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1140; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGACGAGGTAAGTCTGAACACCTTTCGCGAAGACCGGCTCCGTCCTGTCG 60
DB 6641 ATGACGAGGTAAGTCTGAACACCTTTCGCGAAGACCGGCTCCGTCCTGTCG 6582
QY 61 TCCGATTAAGCGCGCATCCGCGACGAACCTGCGAACCCAACTTCAACAAATCACACGTCA 120
DB 6581 TCCGATTAAGCGCGCATCCGCGACGAACCTGCGAACCCAACTTCAACAAATCACACGTCA 6522
QY 121 GAAACATTAAAGCCCTGTGGCCCGCGGATTTCTGGGGTGAACGCTGTGACATTTGTCAC 180
DB 6521 GAAACATTAAAGCCCTGTGGCCCGCGGATTTCTGGGGTGAACGCTGTGACATTTGTCAC 6462
QY 181 GGTCCGAAAGGTTCTGATTCGAGAGGCTTCTTCTCGCGAAGGGCTGATCTGCCCC 240

```

Db      6461 GGTCCGAAAGGTTCTGATTCGAGGCGTTCCGTTCCGCGAAAGGCGTCGATCTGCCCC 6402
QY      241 TCGTGGCGGGAAGAGTGGTGCACATCGTGCAGAGAAATTTCTCAATGTTGTCTCAT 300
Db      6401 TCGTGGCGGGAAGAGTGGTGCACATCGTGCAGAGAAATTTCTCAATGTTGTCTCAT 6342
QY      301 CAACCTGGGAGCTGGATCTGTGGATGGTGAAGATGACATGCGGCAATCACTGGTCTG 360
Db      6341 CAACCTGGGAGCTGGATCTGTGGATGGTGAAGATGACATGCGGCAATCACTGGTCTG 6282
QY      361 CGGCTCCACGACCTATGACATGACCTTTGCGAGCGCTGGAAAGCTGCGACCAACGGTCT 420
Db      6281 CGGCTCCACGACCTATGACATGACCTTTGCGAGCGCTGGAAAGCTGCGACCAACGGTCT 6222
QY      421 CGTTGGCGTACCGAAAGTGAATGTAAGCGCTGCGACGATACGTCGCGCTGTGAATC 480
Db      6221 CGTTGGCGTACCGAAAGTGAATGTAAGCGCTGCGACGATACGTCGCGCTGTGAATC 6162
QY      481 ACTCCGGAAGAAAGCGGCTGGGACGTCGACGTTCCAGCGCTACTCATGTTCAAGTGTGAC 540
Db      6161 ACTCCGGAAGAAAGCGGCTGGGACGTCGACGTTCCAGCGCTACTCATGTTCAAGTGTGAC 6102
QY      541 GTGAGTGAAGACATCCTCGAATCCTTCTCGATGCGATGTTGATGCGTGCATTCCAA 600
Db      6101 GTGAGTGAAGACATCCTCGAATCCTTCTCGATGCGATGTTGATGCGTGCATTCCAA 6042
QY      601 CTCGTATCTCTGGGATTTCTCGCGCCACTACGTAATCGGCTGTCTCATGTAGCAAG 660
Db      6041 CTCGTATCTCTGGGATTTCTCGCGCCACTACGTAATCGGCTGTCTCATGTAGCAAG 5982
QY      661 ATCCGCGGTGAAGCTGATCAAGTTCGCTGCGTATCTGACCAAAATTCATCTGGCGTT 720
Db      5981 ATCCGCGGTGAAGCTGATCAAGTTCGCTGCGTATCTGACCAAAATTCATCTGGCGTT 5922
QY      721 GGTATGAGAGTTGTTAGTGGCGACGGAAGAAAGTGTGACATGCGCAACCGTGCACCTG 780
Db      5921 GGTATGAGAGTTGTTAGTGGCGACGGAAGAAAGTGTGACATGCGCAACCGTGCACCTG 5862
QY      781 GAAATCGCTGTGATGACGTGGCGGCGGATTCACAAAGCGTTGGAACGTGTGGCGAGATTT 840
Db      5861 GAAATCGCTGTGATGACGTGGCGGCGGATTCACAAAGCGTTGGAACGTGTGGCGAGATTT 5802
QY      841 GAGTTTGTTCATGAGGAGTCTGGGGAATCGCGTGGTCCCCGTGATGTTGTCGCCAGCT 900
Db      5801 GAGTTTGTTCATGAGGAGTCTGGGGAATCGCGTGGTCCCCGTGATGTTGTCGCCAGCT 5742
QY      901 GGTCTTGGGCGAGAACTTAACAGATGCTCAGATGTTGAGCAGAGAAATCTGCCCGGTC 960
Db      5741 GGTCTTGGGCGAGAACTTAACAGATGCTCAGATGTTGAGCAGAGAAATCTGCCCGGTC 5682
QY      961 ATGGTTGCGATCAATTCGCGCGCGATCGTGAATGATGATTCGACCTTGCGCCTTAAGTC 1020
Db      5681 ATGGTTGCGATCAATTCGCGCGCGATCGTGAATGATGATTCGACCTTGCGCCTTAAGTC 5622
QY      1021 TTCCGGGAGATCTCGGAGCTCGTCGAAAGCTGGGCGGCACTTTGGGAAATCTTGTGATC 1080
Db      5621 TTCCGGGAGATCTCGGAGCTCGTCGAAAGCTGGGCGGCACTTTGGGAAATCTTGTGATC 5562
QY      1081 TTGCATTATGCAATTTGCCCGCAGCGAGTGTGCGGCGCCCGAATAATATCGGTTGCAAGTA 1140
Db      5561 TTGCATTATGCAATTTGCCCGCAGCGAGTGTGCGGCGCCCGAATAATATCGGTTGCAAGTA 5502

RESULT 9
US-10-007-527a-19/c
; Sequence 19, Application US/10007527A
; Publication No. US20030044807A1
; GENERAL INFORMATION:
; APPLICANT: Tomb, Jean-Francois
; APPLICANT: Bramucci, Michael G.
; APPLICANT: Cheng, Qiong
; APPLICANT: Kostichka, Kristy N.
; TITLE OF INVENTION: Rhodococcus Cloning and Expression Vectors

```

```

; FILE REFERENCE: CL1709 US NA
; CURRENT APPLICATION NUMBER: US/10/007,527A
; CURRENT FILING DATE: 2001-12-05
; PRIOR APPLICATION NUMBER: 60/254,868
; PRIOR FILING DATE: 2000-12-12
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: Microsoft Office 97
; SEQ ID NO: 19
; LENGTH: 20
; TYPE: DNA
; ORGANISM: Primer
US-10-007-527a-19

Query Match
Best Local Similarity 100.0%; Pred. No. 4;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1119 GATTAATTCGGTTCGCAAGT 1138
Db      20 GATTAATTCGGTTCGCAAGT 1

RESULT 10
US-10-007-452-19/c
; Sequence 19, Application US/10007452
; Publication No. US2003003701A1
; GENERAL INFORMATION:
; APPLICANT: Tomb, Jean-Francois
; APPLICANT: Bramucci, Michael G.
; APPLICANT: Cheng, Qiong
; APPLICANT: Kostichka, Kristy N.
; TITLE OF INVENTION: Rhodococcus Cloning and Expression Vectors
; FILE REFERENCE: CL1709 US NA
; CURRENT APPLICATION NUMBER: US/10/007,452
; CURRENT FILING DATE: 2001-11-08
; PRIOR APPLICATION NUMBER: 60/254,868
; PRIOR FILING DATE: 2000-12-12
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: Microsoft Office 97
; SEQ ID NO: 19
; LENGTH: 20
; TYPE: DNA
; ORGANISM: Primer
US-10-007-452-19

Query Match
Best Local Similarity 100.0%; Pred. No. 4;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1119 GATTAATTCGGTTCGCAAGT 1138
Db      20 GATTAATTCGGTTCGCAAGT 1

RESULT 11
US-10-027-632-94798
; Sequence 94798, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; APPLICANT: Ident. Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24

```

```
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 94798
; LENGTH: 454
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-94798
```

```
Query Match          1.8%; Score 20; DB 12; Length 454;
Best Local Similarity 100.0%; Pred. No. 3.6;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      324 GATGTGACGATGACCATGC 343
Db      266 GATGTGACGATGACCATGC 285
```

```
RESULT 12
US-10-027-632-94798
; Sequence 94798, Application US/10027632
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; PRIOR FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 94798
; LENGTH: 454
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-94798
```

```
Query Match          1.8%; Score 20; DB 13; Length 454;
Best Local Similarity 100.0%; Pred. No. 3.6;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      324 GATGTGACGATGACCATGC 343
Db      266 GATGTGACGATGACCATGC 285
```

```
RESULT 13
US-10-225-810-5/c
; Sequence 5, Application US/10225810
; Publication No. US20030157512A1
; GENERAL INFORMATION:
; APPLICANT: Bermingham, Jr., John R.
; TITLE OF INVENTION: Transdorins and Methods of Using Transdorin
; FILE REFERENCE: McLaugh-07165
; CURRENT APPLICATION NUMBER: US/10/225,810
```

```
; CURRENT FILING DATE: 2002-08-21
; NUMBER OF SEQ ID NOS: 76
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 5
; LENGTH: 2055
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-225-810-5
```

```
Query Match          1.8%; Score 20; DB 12; Length 2055;
Best Local Similarity 100.0%; Pred. No. 3.5;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      324 GATGTGACGATGACCATGC 343
Db      1054 GATGTGACGATGACCATGC 1035
```

```
RESULT 14
US-09-805-456-1/c
; Sequence 1, Application US/09805456
; Publication No. US20030170778A1
; GENERAL INFORMATION:
; APPLICANT: Wei, Ming-Hui
; TITLE OF INVENTION: ISOLATED HUMAN TRANSPORTER PROTEINS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN TRANSPORTER PROTEINS,
; FILE REFERENCE: CLO01062
; CURRENT APPLICATION NUMBER: US/09/805,456
; CURRENT FILING DATE: 2001-03-14
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 2093
; TYPE: DNA
; ORGANISM: Human
US-09-805-456-1
```

```
Query Match          1.8%; Score 20; DB 12; Length 2093;
Best Local Similarity 100.0%; Pred. No. 3.5;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      324 GATGTGACGATGACCATGC 343
Db      1010 GATGTGACGATGACCATGC 991
```

```
RESULT 15
US-09-805-456-3/c
; Sequence 3, Application US/09805456
; Publication No. US20030170778A1
; GENERAL INFORMATION:
; APPLICANT: Wei, Ming-Hui
; TITLE OF INVENTION: ISOLATED HUMAN TRANSPORTER PROTEINS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN TRANSPORTER PROTEINS,
; FILE REFERENCE: CLO01062
; CURRENT APPLICATION NUMBER: US/09/805,456
; CURRENT FILING DATE: 2001-03-14
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 46649
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(46649)
; OTHER INFORMATION: n = A,T,C or G
US-09-805-456-3
```

```
Query Match          1.8%; Score 20; DB 12; Length 46649;
Best Local Similarity 100.0%; Pred. No. 3.2;
```

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 324 GATGTGACGATGACCATGC 343

Db 32694 GATGTGACGATGACCATGC 32675

Search completed: November 8, 2003, 23:09:19
Job time : 306 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 8, 2003, 21:28:22 ; Search time 1954 Seconds
(without alignments)

14179.684 Million cell updates/sec

Title: US-10-007-527a-1

Perfect score: 1140

Sequence: 1 atgaccagtgtaagtcgtga.....taatatcggttcgaagta 1140

Scoring table: OLIGO_NUC

Gapop 60.0 , Gapext 60.0

Searched: 22781392 seqs, 12152238056 residues

Word size : 0

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estnu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_hiv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rtd:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	20	1.8	494	9	A1953890 wx70g08.x
2	20	1.8	538	28	AQ392965 CITR1-E1
3	20	1.8	609	14	CA732898 wipic.pk0
4	20	1.8	632	12	BM696484 UI-E-DWO-

5	20	1.8	657	10	BE149632	BE149632 RCL-HT025
6	20	1.8	764	29	B2441145	B2441145 BOND55TR
7	20	1.8	767	28	BH606297	BH606297 BOG027TF
8	20	1.8	781	28	BH716628	BH716628 BOHW23TR
9	20	1.8	790	9	AL043182	AL043182 DKFZP434G
10	20	1.8	820	29	B2447351	B2447351 BONFN04TF
11	20	1.8	858	28	BH079737	BH079737 RPT-24-3
12	20	1.8	1063	28	BH722906	BH722906 BOWM275TF
13	20	1.8	1315	29	B2576656	B2576656 mbr2_504
14	19	1.7	340	29	CC426700	CC426700 PUFRR89TD
15	19	1.7	376	9	AV697120	AV697120 AV697120
16	19	1.7	390	9	AA459477	AA459477 aa27b05.r
17	19	1.7	415	14	N40228	N40228 yx99b08.r1
18	19	1.7	422	9	AW778576	AW778576 EK51911.Y
19	19	1.7	423	9	AW732682	AW732682 db11b07.Y
20	19	1.7	423	14	H85243	H85243 yv85d07.r1
21	19	1.7	452	9	AA201616	AA201616 LD04711.5
22	19	1.7	462	9	AJ318275	AJ318275 AJ318275
23	19	1.7	470	9	AJ412271	AJ412271 AJ412271
24	19	1.7	475	9	AJ318276	AJ318276 AJ318276
25	19	1.7	552	13	BQ528871	BQ528871 3524_1_42
26	19	1.7	560	13	BQ485604	BQ485604 3524_1_16
27	19	1.7	608	28	BH736326	BH736326 BOH094TF
28	19	1.7	630	9	AA193531	AA193531 zt41a11.r
29	19	1.7	643	14	CD224625	CD224625 CCG1_35.C
30	19	1.7	656	12	BM679842	BM679842 UI-E-EJO-
31	19	1.7	667	28	AQ566635	AQ566635 HS_2105_B
32	19	1.7	834	29	B2451465	B2451465 BONC157TF
33	19	1.7	844	29	CC011138	CC011138 PUEB72TD
34	19	1.7	870	29	B2746327	B2746327 PUDJA49TD
35	19	1.7	947	10	BF239778	BF239778 601906381
36	19	1.7	1013	29	CC283164	CC283164 CH261-138
37	19	1.7	1039	29	CNS02W09	AL216450 Tetradon
38	19	1.7	1109	13	BQ276654	BQ276654 AGENCOURT
39	18	1.6	173	14	CB858146	CB858146 RX85 Medi
40	18	1.6	242	9	AA171157	AA171157 ms43c02.r
41	18	1.6	290	10	BA415255	BA415255 BA415255
42	18	1.6	307	10	BB381929	BB381929 BB381929
43	18	1.6	320	10	BB752403	BB752403 BB752403
44	18	1.6	332	9	AI490628	AI490628 EST249182
45	18	1.6	370	10	BF595561	BF595561 su66f04.Y

ALIGNMENTS

RESULT 1
A1953890
LOCUS
DEFINITION
wx70g08.x1 NCI CGAP Brn53 Homo sapiens cDNA clone IMAGE:2549054 3'
similar to SW:YXK1 YEAST P50944 HYPOHETICAL. 80.0 KD PROTEIN IN
POLI-RAS2 INTERGENIC REGION. ; mRNA sequence.

ACCESSION
A1953890
VERSION
A1953890.1 GI:5746200
KEYWORDS
EST.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens

REFERENCE
1 (bases 1 to 494)
AUTHORS
NCI/NINDS-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute / National Institute of Neurological
Disorders and Stroke, Brain Tumor Genome Anatomy Project
(CGAP/BRGAP), Tumor Gene Index

JOURNAL
COMMENT
Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: Chris Moskalko, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D. cDNA Library Preparation: Life
Technologies, Inc. cDNA Library Arrayed by: Christa Prange, The
I.M.A.G.E. Consortium DNA Sequencing by: Washington University
Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
 Insert Length: 1564 Std Error: 0.00
 Seq primer: -40UP from Gibco
 High quality sequence sc0p: 368.

FEATURES

source

1..494

/organism="Homo sapiens"
 /mol_type="rRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:2549054"
 /tissue_type="three pooled meningiomas"
 /lab_host="DH10B"
 /clone_1lb="NCI CGAP Brn53"
 /note="Organ: brain; Vector: pCMV-SPORT6; Site_1: SalI;
 Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT.
 library constructed by Life Technologies."
 BASE COUNT 139 a 108 c 118 g 129 t
 ORIGIN

Query Match 1.8%; Score 20; DB 9; Length 494;
 Best Local Similarity 100.0%; Pred. No. 27;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 324 GATGTTGACGATGACCATGC 343
 |||||
 Db 372 GATGTTGACGATGACCATGC 391

RESULT 2

AO392965/c

LOCUS CITBI-EI-2544L11.TR CITBI-EI Homo sapiens genomic clone 2544L11,
 DEFINITION genomic survey sequence.

ACCESSION

AO392965

AO392965.1 GI:4363988

KEYWORDS

SOURCE

ORGANISM

Homo sapiens (human)

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 538)

Zhao,S., Adams,M.D., Nierman,W., Malek,J., Shizuya,H., Simon,M. and

Venter,J.C.

TITLE

JOURNAL

COMMENT

Use of BAC End Sequences from Caltech Libraries for Sequence-Ready
 Map Building
 Unpublished
 Other GSSs: CITBI-EI-2544L11.TF
 Contact: Shaying Zhao, William Nierman, Mark Adams
 Department of Bukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850
 Tel: 301 838 0200
 Fax: 301 838 0208
 Email: hbe@tigr.org
 Clones are available from Research Genetics (info@resgen.com). BAC
 end search page:
http://www.tigr.org/cdb/nunngen/bac_end_search/bac_end_search.html.
 Seq primer: M13 Reverse
 Class: BAC ends.

FEATURES
 source
 location/Qualifiers

1..538

/organism="Homo sapiens"

/mol_type="genomic DNA"

/db_xref="taxon:9606"

/clone="2544L11"

/sex="male"

/cell_type="sperm"

/clone_1lb="CITBI-EI"

/note="Vector: pBeloBAC11; Site_1: EcoRI; Site_2: EcoRI;
 Caltech Human BAC Library D"

BASE COUNT

ORIGIN

115 a 148 c 113 g 162 t

Query Match 1.8%; Score 20; DB 28; Length 538;
 Best Local Similarity 100.0%; Pred. No. 28;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 324 GATGTTGACGATGACCATGC 343
 |||||
 Db 205 GATGTTGACGATGACCATGC 186

RESULT 3

CA732898/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Triticum aestivum (bread wheat)

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae

1 (bases 1 to 609)

Tingey,S.V., Powell,W., Wolters,P., Dolan,M., Hainey,C., Yuan,Z.,

Miao,G., Caraher,N. and Hanafey,M.K.

DuPont Wheat cDNA Sequence

Unpublished

CONTACT: Scott V. Tingey

Crop Genetics

E. I. DuPont de Nemours and Company

1 Innovation Way, P.O. Box 6104, Newark, DE 19714-6104, USA

Tel: 302-631-2602

Fax: 302-631-2607

Email: Scott.V.Tingey@USA.dupont.com

Seq primer: M13.

FEATURES

source

1..609

/organism="Triticum aestivum"

/mol_type="rRNA"

/db_xref="taxon:4565"

/clone="wlpic.pk006.a5"

/tissue_type="lemma and palea"

/lab_host="DH10B"

/clone_1lb="wlpic"

/note="Vector: pBluescript SK+; Site_1: EcoRI; Site_2:
 XhoI; wheat (Triticum aestivum, H1 Line) lemma and palea"

XhoI; wheat (Triticum aestivum, H1 Line) lemma and palea"

BASE COUNT

95 a 191 c 185 g 133 t

Query Match 1.8%; Score 20; DB 14; Length 609;
 Best Local Similarity 100.0%; Pred. No. 28;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 323 CGATGTTGACGATGACCATG 342
 |||||
 Db 457 CGATGTTGACGATGACCATG 438

RESULT 4

BM696484/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Homo sapiens (human)

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 632)

Bonaldo,M.F., Lennon,G. and Soares,M.B.

REFERENCE

AUTHORS

115 a 148 c 113 g 162 t

TITLE Normalization and subtraction: two approaches to facilitate gene discovery

JOURNAL Genome Res. 6 (9), 791-806 (1996)

MEDLINE 97044477

PUBMED 8889548

COMMENT Contact: Soares, MB
Coordinated Laboratory for Computational Genomics
University of Iowa
375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: bento-soares@iowa.edu

FEATURES Tissue Procurement: Dr. Gregg Hageman
CDNA Library preparation: Dr. M. Bento Soares, University of Iowa
CDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Researchers may obtain clones from Research Genetics (www.reagen.com).
Seq primer: M13 Reverse.
Location/Qualifiers
1..632
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="UI-E-DW0-agj-n-10-0-UI"
/issue_type="1ens"
/dev_stage="adult"
/lab_host="DH10B (Life Technologies) (T1 phage resistant)"
/clone_lib="UI-E-DW0"
/note="Organ: eye; Vector: pT7T3-Pac (Pharmacia) with a modified polylinker; site_1: EcoR I; site_2: Not I; UI-E-DW0 is a CDNA library containing the following tissue(s): lens. The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was ligated to an EcoR I adaptor, digested with Not I, and cloned directionally into pT7T3-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dT)18 tail. The sequence tag for this library is CAGTTAGCGA. This library was created for the program, Gene Discovery in the Visual System, supported by National Eye Institute (NEI)."

BASE COUNT 123 a 196 c 141 g 172 t

ORIGIN

Query Match 1.8%; Score 20; DB 12; Length 632;
Best Local Similarity 100.0%; Pred.No. 28;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 324 GATGTCGACGATGACCATGC 343
|||||
370 GATGTCGACGATGACCATGC 351

RESULT 5 BE149632 657 bp mRNA linear EST 21-JUN-2000
LOCUS BE149632
DEFINITION R01-HT0256-280300-017-d12 HT0256 Homo sapiens cDNA, mRNA sequence.
ACCESSION BE149632
VERSION BE149632.1 GI:8612353
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens

REFERENCE Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 657)
Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R., Nagai, M.A., da Silva, W.Jr., Zago, M.A., Bordin, S., Costa, F.F., Goldman, G.H., Carvalho, A.F., Matsukuma, A., Bala, G.S., Simpson, D.H., Brunstein, A., de Oliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and

TITLE Simpson, A.J.
Shotgun sequencing of the human transcriptome with ORF expressed sequence tags

JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

MEDLINE 20202663

PUBMED 10737800

COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001

FEATURES Email: asimpson@ludwig.org.br
This sequence was derived from the PABSP/ICR Human Cancer Genome Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?l=et2=R01-HT0256-280300-017-d12&f3=2000-03-28&f4=1)
Seq primer: puc 18 forward
High quality sequence start: 25
High quality sequence stop: 637.
Location/Qualifiers
1..657
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/dev_stage="adult"
/clone_lib="HT0256"
/note="Organ: head neck; Vector: puc18; site_1: SmaI; site_2: SmaI; A mini-library was made by cloning products derived from ORFESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

BASE COUNT 183 a 144 c 182 g 148 t

ORIGIN

Query Match 1.8%; Score 20; DB 10; Length 657;
Best Local Similarity 100.0%; Pred.No. 29;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 324 GATGTCGACGATGACCATGC 343
|||||
230 GATGTCGACGATGACCATGC 249

RESULT 6 BZ441145 764 bp DNA linear GSS 13-DEC-2002
LOCUS BZ441145
DEFINITION BOND55TR_BO_1.6_2_KB_tot Brassica oleracea genomic clone BOND55, genomic survey sequence.
ACCESSION BZ441145
VERSION BZ441145.1 GI:26698749
KEYWORDS GSS.
SOURCE Brassica oleracea
ORGANISM Brassica oleracea

REFERENCE Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
1 (bases 1 to 764)
Town, C.D., Van Aken, S., Uterback, T., Koo, H. and Fraser, C.M.
Whole genome shotgun sequencing of Brassica oleracea
Unpublished
Other GSSs: BOND55TF
Contact: Chris Town
TIGR
5712 Medical Center Drive, Rockville, MD 20850, USA.
Tel: 301-838-3523
Fax: 301-838-0208
Email: cdtown@tigr.org
DNA is from a doubled haploid provided by Tom Osborn.
Seq primer: TR

Class: sheared ends.
 FEATURES Location/Qualifiers
 source 1..764
 /organism="Brassica oleracea"
 /mol_type="genomic DNA"
 /strain="TO1000DH3"
 /db_xref="taxon:3712"
 /clone="BONDR55"
 /clone_1b="BO_1.6.2 KB tot"
 /note="Vector: pHO51; Site 1: BstXI; 1.6-2 kb sheared total DNA inserted into pHO51 using BstXI linkers"
 BASE COUNT 246 a 136 c 150 g 232 t
 ORIGIN
 Query Match 1.8%; Score 20; DB 29; Length 764;
 Best Local Similarity 100.0%; Pred. No. 29;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 98 AACTTCACAAATCACCACG 117
 |||||
 DB 423 AACTTCACAAATCACCACG 442
 |||||
 RESULT 7
 LOCUS BH606297 767 bp DNA linear GSS 15-DEC-2001
 DEFINITION BOGJG27F BOGJ Brassica oleracea genomic clone BOGJG27, genomic survey sequence.
 ACCESSION BH606297
 VERSION BH606297.1 GI:17858743
 KEYWORDS GSS.
 SOURCE Brassica oleracea
 ORGANISM Brassica oleracea
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
 ; eurosid II; Brassicales; Brassicaceae; Brassica.
 1 (bases 1 to 767)
 TOWN,C.D., Van Aken,S., Uteback,T., Koo,H. and Fraser,C.M.
 AUTHORS Whole genome shotgun sequencing of Brassica oleracea
 TITLE Unpublished
 JOURNAL Other GSSs: BOGJG27TR
 COMMENT Contact: Chris Town
 TIGR
 9712 Medical Center Drive, Rockville, MD 20850, USA.
 Tel: 301-838-3523
 Fax: 301-838-0208
 Email: cdtown@tigr.org
 DNA is from a doubled haploid provided by Tom Osborn.
 Seq primer: TP
 Class: sheared ends.
 FEATURES Location/Qualifiers
 source 1..767
 /organism="Brassica oleracea"
 /mol_type="genomic DNA"
 /strain="TO1000DH3"
 /db_xref="taxon:3712"
 /clone="BOGJG27"
 /clone_1b="BOGJ"
 /note="Vector: pHO51; Site 1: BstXI; 2-3 kb sheared genomic DNA inserted into pHO51 using BstXI linkers"
 BASE COUNT 192 a 187 c 188 g 200 t
 ORIGIN
 Query Match 1.8%; Score 20; DB 28; Length 767;
 Best Local Similarity 100.0%; Pred. No. 29;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 304 CTCGGACTGATCTGTGC 323
 |||||
 DB 346 CTCGGACTGATCTGTGC 365
 |||||
 RESULT 8

BH716628
 LOCUS BH716628 781 bp DNA linear GSS 20-FEB-2002
 DEFINITION BOHW23TR BO_2.3 KB Brassica oleracea genomic clone BOHW23, genomic survey sequence.
 ACCESSION BH716628
 VERSION BH716628.1 GI:18812783
 KEYWORDS GSS.
 SOURCE Brassica oleracea
 ORGANISM Brassica oleracea
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
 ; eurosid II; Brassicales; Brassicaceae; Brassica.
 1 (bases 1 to 781)
 TOWN,C.D., Van Aken,S., Uteback,T., Koo,H. and Fraser,C.M.
 AUTHORS Whole genome shotgun sequencing of Brassica oleracea
 TITLE Unpublished
 JOURNAL Other GSSs: BOHW23TR
 COMMENT Contact: Chris Town
 TIGR
 9712 Medical Center Drive, Rockville, MD 20850, USA.
 Tel: 301-838-3523
 Fax: 301-838-0208
 Email: cdtown@tigr.org
 DNA is from a doubled haploid provided by Tom Osborn.
 Seq primer: TR
 Class: sheared ends.
 FEATURES Location/Qualifiers
 source 1..781
 /organism="Brassica oleracea"
 /mol_type="genomic DNA"
 /strain="TO1000DH3"
 /db_xref="taxon:3712"
 /clone="BOHW23"
 /clone_1b="BO_2.3 KB"
 /note="Vector: pHO51; Site 1: BstXI; 2-3 kb sheared genomic DNA inserted into pHO51 using BstXI linkers"
 BASE COUNT 218 a 200 c 160 g 203 t
 ORIGIN
 Query Match 1.8%; Score 20; DB 28; Length 781;
 Best Local Similarity 100.0%; Pred. No. 30;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 304 CTCGGACTGATCTGTGC 323
 |||||
 DB 663 CTCGGACTGATCTGTGC 682
 |||||
 RESULT 9
 LOCUS AL043182/c 790 bp mRNA linear EST 29-FEB-2000
 DEFINITION DKFZP434G1123.r1 434 (synonym: htes3) Homo sapiens cDNA clone DKFZP434G1123.5', mRNA sequence.
 ACCESSION AL043182
 VERSION AL043182.1 GI:5422591
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 1 (bases 1 to 790)
 Blum,H., Bauersachs,S., Mewes,H.W., Gassenhuber,J. and Wiemann,S.
 AUTHORS EST (Blum, et al.)
 TITLE Unpublished
 JOURNAL Contact: Blum H
 COMMENT MIPs
 Am Klopferpitz 18a D-82152 Martinsried, Germany
 This is the 5' sequence of the clone insert
 Clone from S. Wiemann, Molecular Genome Analysis, German Cancer Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
 sequenced by IMU (Ludwig Maximilians University,
 Munich/Germany) within the cDNA sequencing consortium of the German Genome Project.

No al sequence available.
This clone (DKFZp434G1123) is available at the RZPD in Berlin.
Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
Berlin-Charlottenburg, GERMANY; Email: clonezrpd.de.

FEATURES

Location/Qualifiers

1..790

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="DKFZp434G1123"

/issue_type="testis"

/dev_stage="adult"

/lab_host="DH10B"

/clone_1lb="434 (synonym: htes3)"

/note="Vector: pSPori1; Site_1: NotI; Site_2: SalI"

BASE COUNT 151 a 249 c 182 g 204 t 4 others

ORIGIN

Query Match 1.8%; Score 20; DB 9; Length 790;

Best Local Similarity 100.0%; Pred. No. 30;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 324 GATGTGACGATGACCATGC 343

Db 470 GATGTGACGATGACCATGC 451

RESULT 10

B2447351

LOCUS B2447351 820 bp DNA linear GSS 13-DEC-2002

DEFINITION BOMF04TR_BO.1.6_2_KB for Brassica oleracea genomic clone BOMF04,

genomic survey sequence.

ACCESSION B2447351

VERSION B2447351.1 GI:26712635

KEYWORDS GSS.

SOURCE Brassica oleracea

ORGANISM

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids

1 (bases 1 to 820)

Town,C.D., Van Aken,S., Utterback,T., Koo,H. and Fraser,C.M.

TITLE

JOURNAL

COMMENT

Unpublished

Other GSSs: BOMF04TR

Contact: Chris Town

TIGR

9712 Medical Center Drive, Rockville, MD 20850, USA.

Tel: 301-838-3523

Fax: 301-838-0208

Email: cdtown@tigr.org

DNA is from a doubled haploid provided by Tom Osborn.

Seq primer: TF

Class: sheared ends.

Location/Qualifiers

1..820

/organism="Brassica oleracea"

/mol_type="genomic DNA"

/strain="TO1000DH3"

/db_xref="taxon:3712"

/clone="BOMF04"

/clone_1lb="BO.1.6_2_KB tot"

/note="Vector: pBOS1; Site 1: BstXI; 1.6-2 kb sheared

total DNA inserted into pBOS1 using BstXI linkers"

BASE COUNT 254 a 170 c 169 g 227 t

ORIGIN

Query Match 1.8%; Score 20; DB 29; Length 820;

Best Local Similarity 100.0%; Pred. No. 30;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 98 AACTGACAAATCACCAGC 117

Db 45 AACTGACAAATCACCAGC 64

RESULT 11

BH079737/c

LOCUS BH079737/c 858 bp DNA linear GSS 16-JUL-2001

DEFINITION RPCI-24-321D14.TV RPCI-24 Mus musculus genomic clone RPCI-24-321D14

, genomic survey sequence.

ACCESSION BH079737

VERSION BH079737.1 GI:14899334

KEYWORDS GSS.

SOURCE Mus musculus

ORGANISM

Mus musculus

(house mouse)

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Unpublished

Other GSSs: RPCI-24-321D14.TV

Contact: Shaying Zhao

Department of Eukaryotic Genomics

The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850, USA

Tel: 301 838 0200

Fax: 301 838 0208

Email: szhao@tigr.org

Clones are derived from the mouse BAC library RPCI-24. For BAC

library availability, please contact Pieter de Jong

(pdejong@mail.cho.org). Clones may be purchased from BACPAC

Resources (<http://www.choi.org/bacpac/orderingframe.htm>). BAC end

page: http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html

Plate: 321 row: D column: 14

Seq primer: T7

Class: BAC ends.

Location/Qualifiers

1..858

/organism="Mus musculus"

/mol_type="genomic DNA"

/strain="C57BL/6J"

/db_xref="taxon:10090"

/clone="RPCI-24-321D14"

/sex="Male"

/cell_type="Spleen/Brain"

/clone_1lb="RPCI-24"

/note="Vector: pTARBAC1; Site 1: BamHI; Site 2: BamHI;

RPCI-24 Mouse BAC library produced by Pieter de Jong. The

library was cloned in the pTARBAC1 cloning vector at the

BamHI sites using MboI partially digested male C57BL/6J

DNA."

BASE COUNT 213 a 185 c 177 g 283 t

ORIGIN

Query Match 1.8%; Score 20; DB 28; Length 858;

Best Local Similarity 100.0%; Pred. No. 30;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 588 GTGAGCTCCAACTCGTAT 607

Db 166 GTGAGCTCCAACTCGTAT 147

RESULT 12

BH722906/c

LOCUS BH722906/c 1063 bp DNA linear GSS 20-FEB-2002

DEFINITION BOMJ275TR_BO.2.3_KB Brassica oleracea genomic clone BOMJ275,

genomic survey sequence.

ACCESSION BH722906

VERSION BH722906.1 GI:18826669

KEYWORDS GSS.

SOURCE Brassica oleracea

```

ORGANISM
Brassica oleracea
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
; eucosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE
1 (bases 1 to 1063)
TOWN,C.D., Van Aken,S., Utebbeck,T., Koo,H. and Fraser,C.M.
TITLE
Whole genome shotgun sequencing of Brassica oleracea
JOURNAL
Unpublished
COMMENT
Other GSSs: BOMJ275TR
Contact: Chris Town
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA.
Tel: 301-838-3523
Fax: 301-838-0208
Email: cdtown@tigr.org
DNA is from a doubled haploid provided by Tom Osborn.
Seq primer: TF
Class: sheared ends.
FEATURES
Location/Qualifiers
source
1..1063
/organism="Brassica oleracea"
/mol_type="genomic DNA"
/strain="TO1000DH3"
/db_xref="taxon:3712"
/clone_1ib="BO_2_3_KB"
/notes="Vector: pHOSt; Site 1: BstXI; 2-3 kb sheared
genomic DNA inserted into pHOSt using BstXI linkers"
BASE COUNT
332 a 227 c 186 g 318 t
ORIGIN
Query Match
Best Local Similarity 100.0%; Score 20; DB 28; Length 1063;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY
98 AACTCAACAATCACCAG 117
|||||
Db
504 AACTCAACAATCACCAG 485

RESULT 13
B2576656 1315 bp DNA linear GSS 17-DEC-2002
LOCUS
msh2_504.y2 msh Pseudomonas aeruginosa genomic clone msh2_504,
DEFINITION
genomic survey sequence.
ACCESSION
B2576656
VERSION
B2576656.1 GI:27211717
KEYWORDS
GSS.
SOURCE
Pseudomonas aeruginosa
ORGANISM
Pseudomonas aeruginosa
Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
Pseudomonadaceae; Pseudomonas.
1 (bases 1 to 1315)
Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Hastings,M.,
Burns,J.L., Kaul,R. and Olsen,M.V.
TITLE
Whole-Genome-Sequence Variation among multiple isolates of
Pseudomonas aeruginosa library
JOURNAL
J. Bacteriol., (2002) In press
COMMENT
Contact: Chris K. Raymond
Genome Center
University of Washington
Box 352145, Seattle, WA 98105-2145, USA
Tel: 2062216954
Fax: 2066857244
Email: craymond@u.washington.edu
Class: shotgun.
FEATURES
Location/Qualifiers
source
1..1315
/organism="Pseudomonas aeruginosa"
/mol_type="genomic DNA"
/strain="M5H"
/db_xref="taxon:287"
/clone="msh2_504"

```

```

/clone_1ib="msh"
/notes="Environmental isolate. Whole genomic shotgun
library."
BASE COUNT
359 a 352 c 221 g 383 t
ORIGIN
Query Match
Best Local Similarity 100.0%; Score 20; DB 29; Length 1315;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY
227 GCTGATCTGCCCTGCTGT 246
|||||
Db
441 GCTGATCTGCCCTGCTGT 460

RESULT 14
CC426700 340 bp DNA linear GSS 19-MAY-2003
LOCUS
PUHER897D.ZM.0.6.1.0_KB_Zea mays genomic clone ZMMBT434010,
DEFINITION
genomic survey sequence.
ACCESSION
CC426700
VERSION
CC426700.1 GI:30906790
KEYWORDS
GSS.
SOURCE
Zea mays
ORGANISM
Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACAD
clade; Panicoideae; Andropogoneae; Zea.
1 (bases 1 to 340)
Whitefaw,C.A., Quackenbush,J., Van Aken,S., Utebbeck,T., Resnick
,A., Fraser,C.M., Yuan,Y., San Miguel,P., Ma,J. and Bennetzen,J.
TITLE
Maize Genomes Consortium
JOURNAL
Unpublished
COMMENT
Contact: Cathy Whitefaw
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-838-5843
Fax: 301-838-0208
Email: whitefaw@tigr.org
Seq primer: TF
Class: sheared ends.
FEATURES
Location/Qualifiers
source
1..340
/organism="Zea mays"
/mol_type="genomic DNA"
/strain="B73"
/db_xref="taxon:4577"
/clone_1ib="ZM_0.6.1.0_KB"
/notes="Vector: pCR4-toFO; Site 1: EcoRI; 0.6-1.0 kb high
cot selected genomic DNA library"
BASE COUNT
118 a 72 c 74 g 76 t
ORIGIN
Query Match
Best Local Similarity 100.0%; Score 19; DB 29; Length 340;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY
1030 ATCTCGAGCTGCTGAAG 1048
|||||
Db
300 ATCTCGAGCTGCTGAAG 318

RESULT 15
AV697120 376 bp mRNA linear EST 16-JAN-2002
LOCUS
AV697120 GKC Homo sapiens cDNA clone GKCEB10 5', mRNA sequence.
DEFINITION
AV697120
ACCESSION
AV697120.1 GI:10298983
VERSION
AV697120.1 GI:10298983
KEYWORDS
EST.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```

REFERENCE
AUTHORS

1 (bases 1 to 376)
Mammalia: Eutheria; Primates; Catarrhini; Homidae; Homo.

TITLE

Insight into hepatocellular carcinogenesis at transcriptome level
by comparing gene expression profiles of hepatocellular carcinoma
with those of corresponding noncancerous liver

JOURNAL
MEDLINE
PUBMED
COMMENT

Proc. Natl. Acad. Sci. U.S.A. 98 (26), 15089-15094 (2001)
21625106
11752456
Contact: Zeguang Han
Chinese National Human Genome Center at Shanghai
351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai
201203, P. R. China
Tel: 86-21-50801919 (ex.45)
Fax: 86-21-50801922
Email: hanzg@chgc.sh.cn
This clone is available at CHGC in Shanghai.

FEATURES

source

1..376
Location/Qualifiers
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="GKCE0E10"
/tissue_type="hepatocellular carcinoma"
/dev_stage="Adult"
/lab_host="SOLR"
/clone_1b="GKCE"
/note="Vector: pBluescript sk(-); Site_1: EcoRI; Site_2:
XhoI"

BASE COUNT 76 a 113 c 113 g 74 t
ORIGIN

Query Match

Best Local Similarity 1.7%; Score 19; DB 9; Length 376;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 227 GCTGATCTGCCCCCTGCTG 245
|||
Db 254 GCTGATCTGCCCCCTGCTG 272

Search completed: November 8, 2003, 23:02:25
Job time : 1958 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 8, 2003, 12:11:30 ; Search time 2940 Seconds
(without alignments)

15862.917 Million cell updates/sec

Title: US-10-007-527A-1

Sequence: 1140

1 atgaccagcgaagcgcgcga.....taatacgggttcgcaagtcga 1140

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 2888711 seqs, 20454813386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : GenBank1:
1: gb_ba:*
2: gb_hcg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: gb_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_hcg_hum:*
31: em_hcg_in:*
32: em_hcg_oher:*
33: em_hcg_mus:*
34: em_hcg_pin:*
35: em_hcg_rod:*
36: em_hcg_mam:*
37: em_hcg_vtl:*
38: em_sy:*
39: em_hcgo_hum:*
40: em_hcgo_mus:*
41: em_hcgo_other:*

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1140	100.0	1140	6	AX548642 Sequence
2	1140	100.0	6334	6	AX178757 Rhodococc
3	1140	100.0	6334	6	AX548646 Sequence
4	1140	100.0	9652	6	AX548648 Sequence
5	1140	100.0	11241	6	AX548647 Sequence
6	1140	100.0	11241	12	AX180162 Shuttle v
7	88.4	8.3	3540	1	AX150274
8	94.8	7.8	2439	1	APU83788
9	55.8	4.9	4603	1	AX172684
10	51.6	4.5	2297	6	E17316
11	51.6	4.5	5750	1	AF085719
12	49	4.3	125020	9	AF429315
13	47.2	4.1	2051	1	PPR6662
14	46.4	4.1	125020	9	AF429315
15	44.4	3.9	2000	6	AX655393
16	42.4	3.7	9367	1	SNR243257
17	41.4	3.6	13651	1	AB011844
18	41.2	3.6	12855	1	AE333885
19	40.6	3.6	1841	6	AX700509
20	40.6	3.6	1841	10	MMPTX3
21	40.6	3.6	163132	2	AC121312
22	39.2	3.4	127764	2	AC144481
23	38.8	3.4	248154	2	AC111496
24	38.8	3.4	249495	2	AC095394
25	38.4	3.4	205178	2	AC132020
26	38.4	3.4	228187	2	AC127107
27	38.4	3.4	304078	2	AC129703
28	38.2	3.4	6758	1	AX172685
29	37.6	3.3	122834	9	AC092898
30	37.4	3.3	2708	6	AX700512
31	37.4	3.3	2708	10	MMU33842
32	37.4	3.3	3147	12	AF092035
33	37.4	3.3	4447	1	AF092037
34	37.4	3.3	4447	6	AR173114
35	37.4	3.3	4447	6	AR173115
36	37.4	3.3	4447	6	BD106940
37	37.4	3.3	4447	6	BD106941
38	37.4	3.3	4447	6	BD139585
39	37.4	3.3	4447	6	BD139586
40	37.4	3.3	4457	1	BDFLASM1
41	37.4	3.3	5135	12	SVL13193
42	37.4	3.3	5216	12	SVL13194
43	37.4	3.3	6241	12	AF092036
44	37.4	3.3	6272	12	AF091268
45	37.4	3.3	6398	12	AF092931

ALIGNMENTS

RESULT 1
AX548642
LOCUS AX548642 1140 bp DNA linear PAT 27-NOV-2002
DEFINITION Sequence 1 from Patent WO02055709.
ACCESSION AX548642
VERSION AX548642.1 GI:25813612
KEYWORDS
SOURCE Rhodococcus erythropolis
ORGANISM Rhodococcus erythropolis
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Corynebacterineae; Nocardiaceae; Rhodococcus.
REFERENCE
1 Bramucci, M.G., Cheng, Q., Kostichka, K.N. and Tomb, J.F.
Rhodococcus cloning and expression vectors
Patent: WO 02055709-A 1 18-JUL-2002;
JOURNAL

Pred. No. is the number of results predicted by chance to have a

FEATURES
 source
 E.I. DU PONT DE NEMOURS AND COMPANY (US)
 Location/Qualifiers
 1..1140
 /organism="Rhodococcus erythropolis"
 /mol_type="genomic DNA"
 /strain="AN12"
 /db_xref="taxon:1833"

BASE COUNT 241 a 287 c 347 g 265 t

ORIGIN

Query Match 100.0%; Score 1140; DB 6; Length 1140;
 Best Local Similarity 100.0%; Pred. No. 5.5e-286;
 Matches 1140; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 ATGACGAGCGTAAGTGTGACACCTTTCCGGCAAAAGACGGGCTCCCGCTCCGCTGTCG 60
Db 1 ATGACGAGCGTAAGTGTGACACCTTTCCGGCAAAAGACGGGCTCCCGCTCCGCTGTCG 60
QY 61 TCCGATTAAGCGCGGATCCGGGACGAACTGCGACCCCAAATTCAACAAATCACACGTCA 120
Db 61 TCCGATTAAGCGCGGATCCGGGACGAACTGCGACCCCAAATTCAACAAATCACACGTCA 120
QY 121 GAAACATTTAAGCGCTGTGGCGCGGCGGATTTCTGGGCTGAACGGTGTACCATTTGTCAAC 180
Db 121 GAAACATTTAAGCGCTGTGGCGCGGCGGATTTCTGGGCTGAACGGTGTACCATTTGTCAAC 180
QY 181 GGTCCGAAAGTTCTGATTCGAGAGGCTTCGTTCCCTGCGAAAGGGCTGATTCGCCCC 240
Db 181 GGTCCGAAAGTTCTGATTCGAGAGGCTTCGTTCCCTGCGAAAGGGCTGATTCGCCCC 240
QY 241 TCCCTGTGCGGAAAGTCCGTGCACTGCTGACAGCAAGAAATTTCTCAAGTTGTGCTCAT 300
Db 241 TCCCTGTGCGGAAAGTCCGTGCACTGCTGACAGCAAGAAATTTCTCAAGTTGTGCTCAT 300
QY 301 CAACTCGGAGCTGGAATCTGTGGGATGTGAGAGAGACATACGCGCATACACTGCTGTGAG 360
Db 301 CAACTCGGAGCTGGAATCTGTGGGATGTGAGAGAGACATACGCGCATACACTGCTGTGAG 360
QY 361 CGGCTCCACGACCTATGACCTGACCTTTGGGACGCTGGAAGAGCTGACCAACGGTCTGT 420
Db 361 CGGCTCCACGACCTATGACCTGACCTTTGGGACGCTGGAAGAGCTGACCAACGGTCTGT 420
QY 421 CGTTGGCGTACGGAAGCTGAAATGTACGGCTGCGACGAGATACGTCGCGCTGTGAAATC 480
Db 421 CGTTGGCGTACGGAAGCTGAAATGTACGGCTGCGACGAGATACGTCGCGCTGTGAAATC 480
QY 481 ACTCAGGAAAAAAGCGGCTGGGACGCTGCAAGTTCAAGCGCTACTCATTTCTGAGTGTGAC 540
Db 481 ACTCAGGAAAAAAGCGGCTGGGACGCTGCAAGTTCAAGCGCTACTCATTTCTGAGTGTGAC 540
QY 541 GTGAGTGAAGACATCTCGAATCTCTCGGATGCGATGTCGATCGGTGACCTTCCAA 600
Db 541 GTGAGTGAAGACATCTCGAATCTCTCGGATGCGATGTCGATCGGTGACCTTCCAA 600
QY 601 CTCGATCTCTGGAATTTGCTGGGCACTACGTAATTCGAGTGTCTCATGTACGAAG 660
Db 601 CTCGATCTCTGGAATTTGCTGGGCACTACGTAATTCGAGTGTCTCATGTACGAAG 660
QY 661 ATGCGGGGGAAGCTGATCAAGTTCTGCGCTGCTATCTACCAAAATTTGCAATTCGGCCT 720
Db 661 ATGCGGGGGAAGCTGATCAAGTTCTGCGCTGCTATCTACCAAAATTTGCAATTCGGCCT 720
QY 721 GGTATGAGGTTGTGATGCGGCGGCAAGAAAGTGTGCAATGCGCAACCGTGCACCTCTG 780
Db 721 GGTATGAGGTTGTGATGCGGCGGCAAGAAAGTGTGCAATGCGCAACCGTGCACCTCTG 780
QY 781 GAAATGCTGTTGATGACGTGGGCGGGGATCCAAAGCGTTGAACTGTGGCGCAATTT 840
Db 781 GAAATGCTGTTGATGACGTGGGCGGGGATCCAAAGCGTTGAACTGTGGCGCAATTT 840
QY 841 GAGTTTGTTCATGAGGAGCGTGGGCAATTCGCTGCTCCCTGAGATTCGCTGCCGAGCT 900
Db 841 GAGTTTGTTCATGAGGAGCGTGGGCAATTCGCTGCTCCCTGAGATTCGCTGCCGAGCT 900
  
```

QY 901 GGTCTTGGGCGAGAACTPAACGATGCTCAGATCTGTGAGCGAGAAATCTGCCCGGTC 960
 Db 901 GGTCTTGGGCGAGAACTPAACGATGCTCAGATCTGTGAGCGAGAAATCTGCCCGGTC 960
 QY 961 ATGTTGGATCATTTCCGGCGCGGATCGTGAATGATTTGAGACTTTGTGCCCTTACGTC 1020
 Db 961 ATGTTGGATCATTTCCGGCGCGGATCGTGAATGATTTGAGACTTTGTGCCCTTACGTC 1020
 QY 1021 TTCCGCGAGATCTCGGACTCGTGAAGCTGGCGGACCTTGGGAAATCTTCGATCAC 1080
 Db 1021 TTCCGCGAGATCTCGGACTCGTGAAGCTGGCGGACCTTGGGAAATCTTCGATCAC 1080
 QY 1081 TTGCAATTCGATTCGCCGAGCGAGATGCGGCCCGCGATPAATTCGTTGCGAAGTGA 1140
 Db 1081 TTGCAATTCGATTCGCCGAGCGAGATGCGGCCCGCGATPAATTCGTTGCGAAGTGA 1140

RESULT 2
 A178757/c 6334 bp DNA linear BCT 01-MAR-2003
 LOCUS
 DEFINITION Rhodococcus erythropolis cryptic plasmid pAN12, complete sequence.
 ACCESSION A178757
 VERSION A178757.1 GI:28628256
 KEYWORDS
 SOURCE
 ORGANISM
 Rhodococcus erythropolis
 Bacteria; Actinobacteridae; Actinomycetales;
 Corynebacterineae; Nocardiaceae; Rhodococcus.
 REFERENCE
 1 (bases 1 to 6334)
 Koestichka, K., Tao, L., Brammucci, M., Tomb, J.-F., Nagarajan, V. and
 Cheng, Q.
 A small cryptic plasmid from Rhodococcus erythropolis:
 Characterization and utility for gene expression
 TITLE
 Unpublished
 JOURNAL
 2 (bases 1 to 6334)
 REFERENCES
 Submitted (11-NOV-2002) CRAD, E.I. Dupont de Nemours Inc.,
 Experimental Station, Wilmington, DE 19880-0326, USA
 AUTHORS
 Cheng, Q.
 JOURNAL
 Direct Submission
 Submitted (11-NOV-2002) CRAD, E.I. Dupont de Nemours Inc.,
 Experimental Station, Wilmington, DE 19880-0326, USA
 TITLE
 Unpublished
 JOURNAL
 2 (bases 1 to 6334)
 REFERENCES
 Submitted (11-NOV-2002) CRAD, E.I. Dupont de Nemours Inc.,
 Experimental Station, Wilmington, DE 19880-0326, USA
 AUTHORS
 Cheng, Q.

FEATURES
 source
 1..6334
 /organism="Rhodococcus erythropolis"
 /mol_type="genomic DNA"
 /db_xref="taxon:1833"
 /plasmid="pAN12"
 /note="cryptic plasmid"
 complement(1912..3051)
 /gene="rep"
 complement(1912..3051)
 /gene="rep"
 /note="Rep"
 /codon_start=1
 /transl_table=11
 /product="replication protein"
 /protein_id="AAO31602.1"
 /db_xref="GI:28628257"
 /translation="MTSYSAEHLGSKDRPVLVSDDKGIHRLPKLOITSETFN
 AGRPISGVNCTIYNGKRGSGFGLRSCGKMTCPCCAGYGAARADEISGVVHOL
 GTGKGVHVAHMLFSGDVSENLTFGLSAWKAATNGRMRTEREMYGCDYVRAVEI
 THGNKGVHVAHMLFSGDVSENLTFGLSAWKAATNGRMRTEREMYGCDYVRAVEI
 RRIIGEAQOVLAAYITKLAGVGVSGDGSGHGRNRPWELIADVAGDPALDEL
 WREFEGSGRRALMSRGLRABAGLGAELTDAQIVOEBSAPVMVAIIPASMMIR
 TCAVPGEIIGLVAAGATWENLRHMLHRLPAAVVRPIISVRK"
 complement(4288..5178)
 /gene="div"
 complement(4288..5178)
 /gene="div"
 complement(4288..5178)
 /note="div"
 /codon_start=1
 /transl_table=11
 /product="putative cell division protein"

```
/protein_id="AA031603.1"
/db_xref="GI:28628258"
/translation="MDQDTIPAIQWNLADPFWVDIAKDAHWLIQKTSKSKSC
TYNLAQASNPVAVRVGVDPTSVLAPEVNRVAPENITELGINDDKYVQJQFKA
ESDRLECFWDRIDIKSLFSPALPLIVLEFPGLIGADQFDNMLKPADRVAP
RTSLIVROIAQASAKAGIRMLILAOBAAISVGNASRNPVMTLRVDEPSVKMLH
PNATRECALVSEFVPGQGFPODPIGRQMRIRVRYGEISTASYENADLAYEALN
IDRAQMTTASEPHLADIG"
BASE COUNT      1386 a 1807 c 1681 g 1460 t
ORIGIN

Query Match      100.0%; Score 1140; DB 1; Length 6334;
Best Local Similarity 100.0%; Pred. No. 5.3e-286;
Matches 1140; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1  ATGACGAGCGTAAAGTGTGAACAACCTTTCGGCAAAAGCCGCGCTCCGCTCTGTGTCG 60
Db      3051 ATGACGAGCGTAAAGTGTGAACAACCTTTCGGCAAAAGCCGCGCTCCGCTCTGTGTCG 2992

QY      61  TCCGATPAAGCGCGGCATCCGGGACGAACTGCGAACCCAACTTCAACAAATCACACGTCG 120
Db      2991 TCCGATPAAGCGCGGCATCCGGGACGAACTGCGAACCCAACTTCAACAAATCACACGTCG 2932

QY      121  GAAACATTTAAAGCCCTGTGCGCGCGCGGATTTCTGCGGTGAAGCGGTGACCAATTCATC 180
Db      2931 GAAACATTTAAAGCCCTGTGCGCGCGCGGATTTCTGCGGTGAAGCGGTGACCAATTCATC 2872

QY      181  GGTCCGAAAGGTTCTGGAATTCGAGAGGCTTCTGCTCCGCGAAAGGCGTGAATCTGCCCC 240
Db      2871 GGTCCGAAAGGTTCTGGAATTCGAGAGGCTTCTGCTCCGCGAAAGGCGTGAATCTGCCCC 2812

QY      241  TGTGTGCGGGGAAAAGTCGGTGCACATCTGTGCAGACGAAATTTCTCAAGTTGTGTCAT 300
Db      2811 TGTGTGCGGGGAAAAGTCGGTGCACATCTGTGCAGACGAAATTTCTCAAGTTGTGTCAT 2752

QY      301  CAACTGGGAGCTGAGTCTGTGCGATGTGACATGACATGCGCCATACACTGCTGTCAG 360
Db      2751 CAACTGGGAGCTGAGTCTGTGCGATGTGACATGACATGCGCCATACACTGCTGTCAG 2692

QY      361  CGGCTCCAGCACTATGACATGACCTTTCGGCGAGCCTTGAAAAGCTGCGACCAACGGTCGT 420
Db      2691 CGGCTCCAGCACTATGACATGACCTTTCGGCGAGCCTTGAAAAGCTGCGACCAACGGTCGT 2632

QY      421  CGTTGGCGTACGGAACGTGAATGTACGCGTGCAGCGAATACGTGCGCGCTGTGAATC 480
Db      2631 CGTTGGCGTACGGAACGTGAATGTACGCGTGCAGCGAATACGTGCGCGCTGTGAATC 2572

QY      481  ACTCAAGGAAAAAAGCGCTGCGACGCTCCAGTTCACGCGCTACTCATGTTCAAGTGTGAC 540
Db      2571 ACTCAAGGAAAAAAGCGCTGCGACGCTCCAGTTCACGCGCTACTCATGTTCAAGTGTGAC 2512

QY      541  GTGAGTGAAGAACATCTCTCGAATCTTCTCGAGTGCAGATGTTCGATCGGTGAGACTTCCAA 600
Db      2511 GTGAGTGAAGAACATCTCTCGAATCTTCTCGAGTGCAGATGTTCGATCGGTGAGACTTCCAA 2452

QY      601  CTGGAATCTCTGGGAATTTCTGCGCACTACGTAATTCGGGTGATCTCGATGTACGAAAG 660
Db      2451 CTGGAATCTCTGGGAATTTCTGCGCACTACGTAATTCGGGTGATCTCGATGTACGAAAG 2392

QY      661  ATGCGGCGTGAACCTGATCAAGTTCGCTGCGGTATCTGACGAAATTTGATCTGCGCTT 720
Db      2391 ATGCGGCGTGAACCTGATCAAGTTCGCTGCGGTATCTGACGAAATTTGATCTGCGCTT 2332

QY      721  GGTATGAGAGTTGGTGTGCGACGGAAGAAAGTGTGACATGAGCAACCGTGAACCTGCG 780
Db      2331 GGTATGAGAGTTGGTGTGCGACGGAAGAAAGTGTGACATGAGCAACCGTGAACCTGCG 2272

QY      781  GAAATGCTGTTGATGACGCGCGGAGATCCAAAGCGTTGGAACTGTGCGACGAAATTT 840
Db      2271 GAAATGCTGTTGATGACGCGCGGAGATCCAAAGCGTTGGAACTGTGCGACGAAATTT 2212

QY      841  GAGTTGGTTCCATGGAGCGTCCGGCAATGCGGTGCTCCGTTGATTTGCGTCCGAGCT 900
Db      841 GAGTTGGTTCCATGGAGCGTCCGGCAATGCGGTGCTCCGTTGATTTGCGTCCGAGCT 900
```

```
Db      2211 GAGTTGGTTCCATGGAGCGTCCGGCAATGCGGTGCTCCGTTGATTTGCGTCCGAGCT 2152

QY      901  GGTCTTTGGGGCAAACTAAACAGATGCTCAGATCGTTGAGCAGAAAGATCTGCCGATC 960
Db      2151 GGTCTTTGGGGCAAACTAAACAGATGCTCAGATCGTTGAGCAGAAAGATCTGCCGATC 2092

QY      961  ATGCTTGGCATATATTCGCGCGGATCGTGAATGATGATTCGCACTTGTGCGCTTACGTC 1020
Db      2091 ATGCTTGGCATATATTCGCGCGGATCGTGAATGATGATTCGCACTTGTGCGCTTACGTC 2032

QY      1021  TTGCGGAGATCTCGGACTCGTGAAGCTGCGCGCACTTGGGAAAATCTGTGATTCAC 1080
Db      2031 TTGCGGAGATCTCGGACTCGTGAAGCTGCGCGCACTTGGGAAAATCTGTGATTCAC 1972

QY      1081  TTGCAATTATCATTTGCGCGACGCGAGATGTCGCGCCCGCATATATTCGTTGCGAAGTGA 1140
Db      1971 TTGCAATTATCATTTGCGCGACGCGAGATGTCGCGCCCGCATATATTCGTTGCGAAGTGA 1912

RESULT 3
AX548646/c
LOCUS      AX548646      6334 bp      DNA      linear      PAT 27-NOV-2002
DEFINITION Sequence 5 from Patent WO02055709.
ACCESSION AX548646
VERSION    AX548646.1 GI:25813614
KEYWORDS
SOURCE
ORGANISM   Rhodococcus erythropolis
            Rhodococcus erythropolis
            Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
            Corynebacterineae; Nocardiaceae; Rhodococcus.

REFERENCE 1
AUTHORS    Stamuel, M.G., Cheng, Q., Kostelka, K.N. and Tomb, J.F.
TITLE       Rhodococcus cloning and expression vectors
            Patent: WO 02055709-A 5 18-JUL-2002;
            E.I. DU PONT DE NEMOURS AND COMPANY (US)

FEATURES
            location/Qualifiers
            source          1..6334
                           /organism="Rhodococcus erythropolis"
                           /mol_type="genomic DNA"
                           /strain="AN12"
                           /db_xref="taxon:1833"

BASE COUNT      1386 a 1807 c 1681 g 1460 t
ORIGIN

Query Match      100.0%; Score 1140; DB 6; Length 6334;
Best Local Similarity 100.0%; Pred. No. 5.3e-286;
Matches 1140; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1  ATGACGAGCGTAAAGTGTGAACAACCTTTCGGCAAAAGCCGCGCTCCGCTCTGTGTCG 60
Db      3051 ATGACGAGCGTAAAGTGTGAACAACCTTTCGGCAAAAGCCGCGCTCCGCTCTGTGTCG 2992

QY      61  TCCGATPAAGCGCGGCATCCGGGACGAACTGCGAACCCAACTTCAACAAATCACACGTCG 120
Db      2991 TCCGATPAAGCGCGGCATCCGGGACGAACTGCGAACCCAACTTCAACAAATCACACGTCG 2932

QY      121  GAAACATTTAAAGCCCTGTGCGCGCGCGGATTTCTGCGGTGAAGCGGTGACCAATTCATC 180
Db      2931 GAAACATTTAAAGCCCTGTGCGCGCGCGGATTTCTGCGGTGAAGCGGTGACCAATTCATC 2872

QY      181  GGTCCGAAAGGTTCTGGAATTCGAGAGGCTTCTGCTCCGCGAAAGGCGTGAATCTGCCCC 240
Db      2871 GGTCCGAAAGGTTCTGGAATTCGAGAGGCTTCTGCTCCGCGAAAGGCGTGAATCTGCCCC 2812

QY      241  TGTGTGCGGGGAAAAGTCGGTGCACATCTGTGCAGACGAAATTTCTCAAGTTGTGTCAT 300
Db      2811 TGTGTGCGGGGAAAAGTCGGTGCACATCTGTGCAGACGAAATTTCTCAAGTTGTGTCAT 2752

QY      301  CAACTGGGAGCTGAGTCTGTGCGATGTGACATGACATGCGCCATACACTGCTGTCAG 360
Db      2751 CAACTGGGAGCTGAGTCTGTGCGATGTGACATGACATGCGCCATACACTGCTGTCAG 2692

QY      361  CGGCTCCAGCACTATGACATGACCTTTCGGCGAGCCTTGAAAAGCTGCGACCAACGGTCGT 420
Db      361 CGGCTCCAGCACTATGACATGACCTTTCGGCGAGCCTTGAAAAGCTGCGACCAACGGTCGT 420
```

Db 2691 CCGCTCCACGACCTATGGAATGGAATTTGGCGGAGCTTGAAAGCTGCGACCAACGGTCTG 2632
QY 421 CGTTGGCGTAAACGGAAGTGAATATGTAAGCTGCGAGGATATGATGCGGCTGTGTAATC 480
Db 2631 CGTTGGCGTAAACGGAAGTGAATATGTAAGCTGCGAGGATATGATGCGGCTGTGTAATC 2572
QY 481 ACTCAGGAAAAAAGCGCTGCGACGCTCAAGCTTCAAGCGCTACTACTGTTCAAGTGTGAC 540
Db 2571 ACTCAGGAAAAAAGCGCTGCGACGCTCAAGCTTCAAGCGCTACTACTGTTCAAGTGTGAC 2512
QY 541 GTGAGTGAAGAACCTCTCGAATCTTCTCGAGTGCATGTTGATGCGTGAATCTTCAAA 600
Db 2511 GTGAGTGAAGAACCTCTCGAATCTTCTCGAGTGCATGTTGATGCGTGAATCTTCAAA 2452
QY 601 CTGCTATCTCTGGGATTTGCTGCGCACTAAGTAACTTGGGTGTCTGATGTAAGAAAG 660
Db 2451 CTGCTATCTCTGGGATTTGCTGCGCACTAAGTAACTTGGGTGTCTGATGTAAGAAAG 2392
QY 661 ATCGGCGGTGAAGCTGATCAAGTCTGCTGCTGATCTGACGAAATTTGATCTGGCTT 720
Db 2391 ATCGGCGGTGAAGCTGATCAAGTCTGCTGCTGATCTGACGAAATTTGATCTGGCTT 2332
QY 721 GGTATGAGAGTGTGATGAGTGCAGCGAAAAAGTGTGACATGGAACCTGTGACCTTGG 2272
Db 2331 GGTATGAGAGTGTGATGAGTGCAGCGAAAAAGTGTGACATGGAACCTGTGACCTTGG 2272
QY 781 GAATGCGCTGTGATGAGTGCAGCGGAGATCCAAAGCGTTGAACTGTGGCGAGATTT 840
Db 2271 GAATGCGCTGTGATGAGTGCAGCGGAGATCCAAAGCGTTGAACTGTGGCGAGATTT 2212
QY 841 GAGTTTGGTTTGAATGAGACGTGCGGCGAATCGCTGTGCTCCGTGATTCGTTGCCGAGCT 900
Db 2211 GAGTTTGGTTTGAATGAGACGTGCGGCGAATCGCTGTGCTCCGTGATTCGTTGCCGAGCT 2152
QY 901 GGTCTTGGGCGAAGAACTAACAGATCTCAATGCTTGAAGCAAGAAATCTGCCCGGCTC 960
Db 2151 GGTCTTGGGCGAAGAACTAACAGATCTCAATGCTTGAAGCAAGAAATCTGCCCGGCTC 2092
QY 961 ATGTTTGGCGATTCATTCGCGGCGCATGCTGATGATGATTTGGGACTTGTGGCCCTTACGTC 1020
Db 2091 ATGTTTGGCGATTCATTCGCGGCGCATGCTGATGATGATTTGGGACTTGTGGCCCTTACGTC 2032
QY 1021 TTGCGGCGAGATTCCTCGAATCTGTCAGTGCAGTGCAGCTTGGGAAATCTTGTGATCAC 1080
Db 2031 TTGCGGCGAGATTCCTCGAATCTGTCAGTGCAGTGCAGCTTGGGAAATCTTGTGATCAC 1972
QY 1081 TTGCTATTATGATTCGCGGCGAGATGTCGCGCCCGCAATATATGCGTTGCGAAGTGA 1140
Db 1971 TTGCTATTATGATTCGCGGCGAGATGTCGCGCCCGCAATATATGCGTTGCGAAGTGA 1912

RESULT 4
AX548648/c AX548648 9652 bp DNA linear PAT 27-NOV-2002
LOCUS Sequence 7 from Patent WO02055709.
DEFINITION AX548648
ACCESSION AX548648
VERSION AX548648.1 GI:25813616
KEYWORDS
SOURCE Shuttle vector pRHR17
ORGANISM Shuttle vector pRHR17
artificial sequences; vectors.

REFERENCE 1
AUTHORS Bramucci, M.G., Cheng, Q., Kostichka, K.N. and Tomb, J.F.
TITLE Rhodococcus cloning and expression vectors
JOURNAL Patent: WO 02055709-A 7 18-JUL-2002;
E.I. DU PONT DE NEMOURS AND COMPANY (US)

FEATURES
SOURCE 1. 9652
Location/Qualifiers
/organism="Shuttle vector pRHR17"
/mol_type="genomic DNA"
/db_xref="taxon:214938"

BASE COUNT 2097 a 2729 c 2559 g 2267 t

ORIGIN

Query Match 100.0%; Score 1140; DB 6; Length 9652;
Best Local Similarity 100.0%; Pred. No. 5.3e-286;
Matches 1140; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGACCAAGCGTAAAGTGTGAAACACTTTCGCGCAAAAGACGGGCTTCCGTCTGTTGCG 60
Db 5052 ATGACCAAGCGTAAAGTGTGAAACACTTTCGCGCAAAAGACGGGCTTCCGTCTGTTGCG 4993
QY 61 TCCGATTAAGCGCGGATCCGCGACGAGACCCCAACTTCAACAAATACCAACGTC 120
Db 4992 TCCGATTAAGCGCGGATCCGCGACGAGACCCCAACTTCAACAAATACCAACGTC 4933
QY 121 GAAACATTTAAGCGCTGTGCGCGCGCGATTTCTGCGGTGAACGATGTGCAATTTGTCAC 180
Db 4932 GAAACATTTAAGCGCTGTGCGCGCGCGCGATTTCTGCGGTGAACGATGTGCAATTTGTCAC 4873
QY 181 GGTTCGAAAGGTTCTGGAATTCGAGAGCTTGTCTCTGCGAAAGGCTGATCTGCCCC 240
Db 4872 GGTTCGAAAGGTTCTGGAATTCGAGAGCTTGTCTCTGCGAAAGGCTGATCTGCCCC 4813
QY 241 TGTCTGCGGGAAGTGGGTGACATCGTGAACGAAATTTCTCAAGTTGTTGTCAT 300
Db 4812 TGTCTGCGGGAAGTGGGTGACATCGTGAACGAAATTTCTCAAGTTGTTGTCAT 4753
QY 301 CAATCGGAGCTGATCTGTTGCGATGATGATGATGATGATGATGATGATGATGATGATG 360
Db 4752 CAATCGGAGCTGATCTGTTGCGATGATGATGATGATGATGATGATGATGATGATGATG 4693
QY 361 CGGCTTCAGACCTATGATCTGATCTGATCTGATCTGATCTGATCTGATCTGATCTGATCT 420
Db 4692 CGGCTTCAGACCTATGATCTGATCTGATCTGATCTGATCTGATCTGATCTGATCTGATCT 4633
QY 421 CGTTGGCGTAAACGGAAGTGAATATGTAAGCTGCGAGGATTTGGGAAATCTTGTGAAATC 480
Db 4632 CGTTGGCGTAAACGGAAGTGAATATGTAAGCTGCGAGGATTTGGGAAATCTTGTGAAATC 4573
QY 481 ACTCAGGAAAAAAGCGCTGCGACGCTCAAGCTTCAAGCTTCAAGCTTCAAGCTTCAAGCTT 540
Db 4572 ACTCAGGAAAAAAGCGCTGCGACGCTCAAGCTTCAAGCTTCAAGCTTCAAGCTTCAAGCTT 4513
QY 541 GTGAGTGAAGAACCTCTCGAATCTTCTCGAGTGCATGTTGATGCGTGAAGCTTTCGAA 600
Db 4512 GTGAGTGAAGAACCTCTCGAATCTTCTCGAGTGCATGTTGATGCGTGAAGCTTTCGAA 4453
QY 601 CTGCTATCTCTGGGATTTGCTGCGCACTACGTAATTCGGGATGTCGATGTAAGAAAG 660
Db 4452 CTGCTATCTCTGGGATTTGCTGCGCACTACGTAATTCGGGATGTCGATGTAAGAAAG 4393
QY 4392 ATCGGCGGTGAAGCTGATCAAGTCTCGCTGCTGATCTGACGAAATTTGATCTGGCGTT 4333
QY 721 GGTATGAGAGTGTGATGAGTGCAGCGAAAAAGTGTGATGATGATGATGATGATGATGATG 780
Db 4332 GGTATGAGAGTGTGATGAGTGCAGCGAAAAAGTGTGATGATGATGATGATGATGATGATG 4273
QY 781 GAAATGCTGTGATGATGAGTGCAGCGGAGATCCAAAGCTTGGAACTGTGGAGAAATTT 840
Db 4272 GAAATGCTGTGATGATGAGTGCAGCGGAGATCCAAAGCTTGGAACTGTGGAGAAATTT 4213
QY 841 GAGTTTGGTTTGAATGAGACGTGCGGCGAATCGCTGTGCTCCGTGATTCGTTGCCGAGCT 900
Db 4212 GAGTTTGGTTTGAATGAGACGTGCGGCGAATCGCTGTGCTCCGTGATTCGTTGCCGAGCT 4153
QY 901 GGTCTTGGGCGAAGAACTAACAGATCTCAATGCTTGAAGCAAGAAATCTGCCCGGCTC 960
Db 4152 GGTCTTGGGCGAAGAACTAACAGATCTCAATGCTTGAAGCAAGAAATCTGCCCGGCTC 4093
QY 961 ATGTTTGGCGATTCATTCGCGGCGCATGCTGATGATGATTTGGGACTTGTGGCCCTTACGTC 1020
Db 4092 ATGTTTGGCGATTCATTCGCGGCGCATGCTGATGATGATTTGGGACTTGTGGCCCTTACGTC 4033

QY 1021 TTCCGGAGATCTCGGACTCGTCAGAGCTGGGCGACATTGGGAAAATCTTGATGATCAC 1080
Db TTCCGGAGATCTCGGACTCGTCAGAGCTGGGCGACATTGGGAAAATCTTGATGATCAC 3973
QY 1081 TTGCATTATGATTTGCCCGAGCGAGTGTGGCGGCCCGCATTAATATCGTTGGCAAGTGA 1140
Db TTGCATTATGATTTGCCCGAGCGAGTGTGGCGGCCCGCATTAATATCGTTGGCAAGTGA 3913

RESULT 5
AX548647/c

LOCUS AX548647 11241 bp DNA linear PAT 27-NOV-2002
DEFINITION Sequence 6 from Patent WO02055709.
ACCESSION AX548647
VERSION AX548647.1 GI:25813615

KEYWORDS
SOURCE Shuttle vector pRHR17
ORGANISM Shuttle vector pRHR17
Artificial sequences; vectors.

REFERENCE

1 Brannucci, M.G., Cheng, Q., Kostichka, K.N. and Tomb, J.F.
AUTHORS Rhodococcus cloning and expression vectors
TITLE Patent: WO 02055709-A 6 18-JUL-2002;
JOURNAL E.I. DU PONT DE NEMOURS AND COMPANY (US)

FEATURES
source Location/Qualifiers

1..11241
/organism="Shuttle vector pRHR17"

/mol_type="genomic DNA"

/db_xref="taxon:214938"

BASE COUNT 2517 a 3109 c 2945 g 2670 t
ORIGIN

Query Match

100.0%; Score 1140; DB 6; Length 11241;

Best Local Similarity 100.0%; Pred. No. 5.2e-286;
Matches 1140; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGACGAGCGTAAGTCTGTAACACCTTTCGCGAAAGACCGGCTCCGTCGTGTGCG 60
Db 6641 ATGACGAGCGTAAGTCTGTAACACCTTTCGCGAAAGACCGGCTCCGTCGTGTGCG 6582
QY 61 TTCCGATTAAGCGGGCATTCGCGGACGAACTGGGACCCAACTTCAAAATCCAGTCA 120
Db 6581 TTCCGATTAAGCGGGCATTCGCGGACGAACTGGGACCCAACTTCAAAATCCAGTCA 6522
QY 121 GAAACATTTAAGCCTGTGGCCCGGCGGATTTCTGGCGTGAAGCGGTGACCATTTGCAAC 180
Db 6521 GAAACATTTAAGCCTGTGGCCCGGCGGATTTCTGGCGTGAAGCGGTGACCATTTGCAAC 6462
QY 181 GGTCCGAAAGGTTCTGGAATTCGAGAGGCTTCTGTCCTGCGAAAGGGCTGGATCTGCCC 240
Db 6461 GGTCCGAAAGGTTCTGGAATTCGAGAGGCTTCTGTCCTGCGAAAGGGCTGGATCTGCCC 6402
QY 241 TGTCTGGCGGAAAAAGTCGGTGCAATCGTGCAAGAAATTTCTCAAGTTGTGTGTCAT 300
Db 6401 TGTCTGGCGGAAAAAGTCGGTGCAATCGTGCAAGAAATTTCTCAAGTTGTGTGTCAT 6342
QY 301 CAACCTGGGAGCTGATCTGTGGATGTGACATGACATGACATGACATGACATGACATGAC 360
Db 6341 CAACCTGGGAGCTGATCTGTGGATGTGACATGACATGACATGACATGACATGACATGAC 6282
QY 361 CGGCTCCAGCACTTATGACTGAGCTTTCGCGACGCTTGAAGACTGCAACCAACGCGTCTG 420
Db 6281 CGGCTCCAGCACTTATGACTGAGCTTTCGCGACGCTTGAAGACTGCAACCAACGCGTCTG 6222
QY 421 CGTTGGCGTACGGAAGTGAATGTACGCTGTGCAAGGATACGTGCGCTGTTGAAATC 480
Db 6221 CGTTGGCGTACGGAAGTGAATGTACGCTGTGCAAGGATACGTGCGCTGTTGAAATC 6162
QY 481 ACTCAAGGAAAAAAGCGCTGGGACGTCACGTTCAAGGCTACTCATGTTCAAGTGGTAC 540
Db 6161 ACTCAAGGAAAAAAGCGCTGGGACGTCACGTTCAAGGCTACTCATGTTCAAGTGGTAC 6102

QY 541 GTGAGTGAACATCTCGTAATCTCTCGATGCGATGCGATCGATCGATCGATCGATCGAA 600
Db 6101 GTGAGTGAACATCTCGTAATCTCTCGATGCGATGCGATCGATCGATCGATCGATCGAA 6042
QY 601 CTCGTATCTCTGGGATTTGCTGCGGACATGATATTCGAGGTTGCTGATGTAAG 660
Db 6041 CTCGTATCTCTGGGATTTGCTGCGGACATGATATTCGAGGTTGCTGATGTAAG 5982
QY 661 ATCGCGGTTGAAGTGTATCAAGTTCTCGTCGATTCGACGAAATTCGATCTGCGCTT 720
Db 5981 ATCGCGGTTGAAGTGTATCAAGTTCTCGTCGATTCGACGAAATTCGATCTGCGCTT 5922
QY 721 GGTATGAGGTTGTGATGAGCGGACGGAAGAAATGTGTGACATGCGAACCTGTCAG 780
Db 5921 GGTATGAGGTTGTGATGAGCGGACGGAAGAAATGTGTGACATGCGAACCTGTCAG 5862
QY 781 GAAATCGCTGTGATGACAGTGGGCGGAGATTCACAAAGCTTGAACCTGAGAGATT 840
Db 5861 GAAATCGCTGTGATGACAGTGGGCGGAGATTCACAAAGCTTGAACCTGAGAGATT 5802
QY 841 GAGTTGGTTGATGAGGACGTCGAGCAATCGCGTGTGCTCGATGATGCGGAGCT 900
Db 5801 GAGTTGGTTGATGAGGACGTCGAGCAATCGCGTGTGCTCGATGATGCGGAGCT 5742
QY 901 GGTCTTTGGGCGAACTAACAATGCTGATGCTGTTGACAGAAATCTGCCCCGCTC 960
Db 5741 GGTCTTTGGGCGAACTAACAATGCTGATGCTGTTGACAGAAATCTGCCCCGCTC 5682
QY 961 ATGTTGCGATCATTCGCGCGGCGGATCGTGGATGATGATTCGGAATTTGCGGCTTAC 1020
Db 5681 ATGTTGCGATCATTCGCGCGGCGGATCGTGGATGATGATTCGGAATTTGCGGCTTAC 5622
QY 1021 TTCCGGAGATCTCGGACTCGTCAGAGCTGGGCGGACCTTGGGAAAATCTTGATGATCAC 1080
Db 5621 TTCCGGAGATCTCGGACTCGTCAGAGCTGGGCGGACCTTGGGAAAATCTTGATGATCAC 5562
QY 1081 TTGCATTATGATTTGCCCGAGCGAGTGTGGCGGCCCGCATTAATATCGTTGGCAAGTGA 1140
Db 5561 TTGCATTATGATTTGCCCGAGCGAGTGTGGCGGCCCGCATTAATATCGTTGGCAAGTGA 5502

RESULT 6
AY180162/c

LOCUS AY180162 11241 bp DNA circular SYN 01-MAY-2003
DEFINITION Shuttle vector pRHR17, complete sequence.
ACCESSION AY180162
VERSION AY180162.1 GI:30313707

KEYWORDS

SOURCE Shuttle vector pRHR17
ORGANISM Shuttle vector pRHR17
Artificial sequences; vectors.

REFERENCE 1 (bases 1 to 11241)
AUTHORS Kostichka, K., Tao, L., Brannucci, M., Tomb, J.-F., Nagarajan, V. and

Cheng, Q.

TITLE A small cryptic plasmid from Rhodococcus erythropolis:
characterization and utility for gene expression

JOURNAL Unpublished
REFERENCE 2 (bases 1 to 11241)
AUTHORS Kostichka, K., Tao, L., Brannucci, M., Tomb, J.-F., Nagarajan, V. and

Cheng, Q.

TITLE Direct Submission
JOURNAL Submitted (14-NOV-2002) CROD, E. I. Dupont de Nemours Inc.,
Experimental Station, Wilmington, DE 19880-0328, USA

FEATURES
Location/Qualifiers

1..11241

/organism="Shuttle vector pRHR17"

/mol_type="genomic DNA"

/db_xref="taxon:214938"

source

/notes="Rhodococcus - Escherichia coli shuttle vector; Sep
I linearize PAN12 cloned into the PvuII site of pBR328"
join(9925..11241,1..3580)
/organism="Cloning vector pBR328"

```
source
    /mol_type="genomic DNA"
    /db_xref="taxon:221946"
    3591. .9924
    /organism="Rhodococcus erythropolis"
    /mol_type="genomic DNA"
    /db_xref="taxon:1833"
    /plasmid="pAN12"
    /note="cryptic plasmid"

BASE COUNT      2517 a      3109 c      2945 g      2670 t

ORIGIN
Query Match      100.0%; Score 1140; DB 12; Length 11241;
Best Local Similarity 100.0%; Pred. No. 5.2e-286;
Matches 1140; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 ATATACCAAGCTAGTGTCTGAACACCTTTCCGGAAGACCGGCTCCGCTCCGCTG 60
Db      6641 ATGACCAAGCTAGTGTCTGAACACCTTTCCGGAAGACCGGCTCCGCTCCGCTG 6582

QY      61 TCCGATTAAGCGCGGCAATCCGCGACGAATGCGACCCAACTTCAACAAATCAACAGTCA 120
Db      6581 TCCGATTAAGCGCGGCAATCCGCGACGAATGCGACCCAACTTCAACAAATCAACAGTCA 6522

QY      121 GAAACATTTAACCGCTGTGGCGCGGCGGATTTCTGCGGTGAACGCGTGTGAACATTGTCAAC 180
Db      6521 GAAACATTTAACCGCTGTGGCGCGGCGGATTTCTGCGGTGAACGCGTGTGAACATTGTCAAC 6462

QY      181 GGTCCGAAGAAGTTCTGATTCGGAAGCCTTCTGCTGCGGAAGAAGGCTGAGATGCCCC 240
Db      6461 GGTCCGAAGAAGTTCTGATTCGGAAGCCTTCTGCTGCGGAAGAAGGCTGAGATGCCCC 6402

QY      241 TGGTGTGCGGGAAGAAGTCGTGACATCGTGACAGCAAAATTTCTGAAGTTGTGCTCAT 300
Db      6401 TGGTGTGCGGGAAGAAGTCGTGACATCGTGACAGCAAAATTTCTGAAGTTGTGCTCAT 6342

QY      301 CAACTCGGGACTGATCTGTTCGATGTGAAGCAATGACATGCGGCAATACAGCTGTGAC 360
Db      6341 CAACTCGGGACTGATCTGTTCGATGTGAAGCAATGACATGCGGCAATACAGCTGTGAC 6282

QY      361 CGGCTCCAGCACTTATGATCTGACATTTCCGCAAGCCTCGGAACCTGGAACCAACGCTGCT 420
Db      6281 CGGCTCCAGCACTTATGATCTGACATTTCCGCAAGCCTCGGAACCTGGAACCAACGCTGCT 6222

QY      421 CGTTGGCGTACGGAACGTGAATGTATCGGCTGACGAGATGAGTGCAGTGAATC 480
Db      6221 CGTTGGCGTACGGAACGTGAATGTATCGGCTGACGAGATGAGTGCAGTGAATC 6162

QY      481 ACTCAGCGAAGAAACCGGCTGCGACGTCACGCTTCAACGCTTCACTCATGTTCAGTGTGAC 540
Db      6161 ACTCAGCGAAGAAACCGGCTGCGACGTCACGCTTCAACGCTTCACTCATGTTCAGTGTGAC 6102

QY      541 GTGAGTGAAGAACTCTTCGAATCCTTCTCGGATGCGGATGTTCCATCGGTGACATTCGCA 600
Db      6101 GTGAGTGAAGAACTCTTCGAATCCTTCTCGGATGCGGATGTTCCATCGGTGACATTCGCA 6042

QY      601 CTGCTATCTCTGGGATTTGCTGCGCACTAGTATTCGGGTGCTCGATGTACGAAG 660
Db      6041 CTGCTATCTCTGGGATTTGCTGCGCACTAGTATTCGGGTGCTCGATGTACGAAG 5982

QY      661 ATCGGCGGTGAAGCTGATCAAGTTCGCTGCGTATCTGACGAAGAAATTCATCTGCGCTT 720
Db      5981 ATCGGCGGTGAAGCTGATCAAGTTCGCTGCGTATCTGACGAAGAAATTCATCTGCGCTT 5922

QY      721 GGTATGAGAGTTGTGTATGTGGGAGGAAAGTGTGACATGCGCAACCGTGCACCTG 780
Db      5921 GGTATGAGAGTTGTGTATGTGGGAGGAAAGTGTGACATGCGCAACCGTGCACCTG 5862

QY      781 GAAATCGCTGTATGACAGTGGCGGAGATCCAAAGGCTGAACTGTGGCGGAATTT 840
Db      5861 GAAATCGCTGTATGACAGTGGCGGAGATCCAAAGGCTGAACTGTGGCGGAATTT 5802

QY      841 GAGTTTGCTTCGATGGAACGTCGCGCAATCGCGTGTCCGCTGATGCTGCTCCGAGCT 900
Db      5801 GAGTTTGCTTCGATGGAACGTCGCGCAATCGCGTGTCCGCTGATGCTGCTCCGAGCT 5742

QY      901 GGTCTTGGGAGAGAACTAAGATGCTCAGATCGTTGAGACAGAAATCTGCCCGGCTC 960
Db      5741 GGTCTTGGGAGAGAACTAAGATGCTCAGATCGTTGAGACAGAAATCTGCCCGGCTC 5682

QY      961 ATGTTTGGCATTCATTCGCGCGGATCGTGAATGATGATTCGACCTTGTCCGCTTACGTC 1020
Db      5681 ATGTTTGGCATTCATTCGCGCGGATCGTGAATGATGATTCGACCTTGTCCGCTTACGTC 5622

QY      1021 TTGCGGAGATCCCTGCGACTGTGCAAGCTGCGCGGACCTTGGGAAATCTCGTATAC 1080
Db      5621 TTGCGGAGATCCCTGCGACTGTGCAAGCTGCGCGGACCTTGGGAAATCTCGTATAC 5562

QY      1081 TTGCAATTCGATTCCTCCGACGAGCGATGTCGCGCCCGATATATTCGTTGCAAGTGA 1140
Db      5561 TTGCAATTCGATTCCTCCGACGAGCGATGTCGCGCCCGATATATTCGTTGCAAGTGA 5502
```

```
RESULT 7
AY150274
LOCUS      3540 bp      DNA      circular BCT 02-JAN-2003
DEFINITION
Proionibacterium granulosum cryptic plasmid pPG01, complete
ACCESSION
AY150274
VERSION
AY150274.1 GI:27465054
KEYWORDS
Proionibacterium granulosum
Proionibacterium granulosum
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Proionibacterineae; Proionibacteriaceae; Proionibacterium.
REFERENCE
1 (bases 1 to 3540)
Farrar, M.D. and Holland, K.T.
Isolation and characterisation of a cryptic plasmid from the human
skin commensal proionibacterium granulosum
Unpublished
2 (bases 1 to 3540)
Farrar, M.D. and Holland, K.T.
Direct Submision
Submitted (13-SEP-2002) Skin Research Centre, Division of
Microbiology, University of Leeds, Leeds LS2 9JT, UK
LOCATION/Qualifiers
1..3540
/organism="Proionibacterium granulosum"
/mol_type="genomic DNA"
/strain="PF283"
/db_xref="taxon:33011"
/plasmid="cryptic plasmid pPG01"
74..280
/codon_start=1
/transl_table=11
/product="orf3"
/protein_id="AA078121.1"
/db_xref="GI:27465055"
/transl_table="MGGQYILNRCYNGSHNEITIGYPMNGHYIYMPDALLSAMPL
SSGFSFGHVEBPRAVEYFHCADHD"
449..11792
/gene="tra"
449..11792
/gene="tra"
/gene="tra"
/note="plasmid transfer protein"
/codon_start=1
/transl_table=11
/product="Tpa"
/protein_id="AA078122.1"
/db_xref="GI:27465056"
/transl_table="MAGERTYVLTGYVMRLIPLVILLGVAFACWCSYHLVYVIGRA
VAALAEAPPSYQVQKMOVPAFYVYRWMSLAAXFIESYGDRIYRPLQMI
RSDRYHLVLRYPAGLADSAALERGAALIORLGRKSMRTCIYVPAHGLDIPR
DARGELBPAPETSMNVYGVKPGSGEVLDSLSPSHLVGAKRSKGSFYVGLL
DQWRHLPLYVAGVDPGILLFNEIAGDGDALRSKLTNDNDAAYOVU.SMTDEND
RTIYLNCERDKMSRNPDESPPGRLLIYLTIEYRQLIRLONPDSARARSDDRRA
SKAAGLVRIAYGATVGVLLVLTQRPDAKTIIGPLRAQLITRVTFAQSDGLRMSH
```

gene	1979. .3280	PELSESEVQKONSMASGVGFLEADVPLTRFRSRYBALTDLHRPGASVQIDLIQ"
CDS	1979. .3280	/gene="rep"
	1979. .3280	/gene="rep"
	1979. .3280	/note="Plasmid replication protein"
	1979. .3280	/codon_start=1
	1979. .3280	/transl_table=1
	1979. .3280	/product="Rep"
	1979. .3280	/protein_id="PANT8123.1"
	1979. .3280	/db_xref="GI:27465057"
	1979. .3280	/translation="MLNIVTLSKAOPHRLCSEHEPCABARSRERORARGLGTTAT
	1979. .3280	FVADTSCRRKREGRHRRHYEMRDLNAPVMPLEFRVKKCAVPSQRIALMAHGAG
	1979. .3280	YAGTALGCGWACVCYCAKISAHRRDLAVVQYAVLGGKVSMLTIQRHHAQDILA
	1979. .3280	ELMALSLOGMNAVTSGRWQEPFCAQLQGVQNAVEYTHSHSHHVAHVLAISKDP
	1979. .3280	TSVDGTLRHRKQGRRTTPEEVRQEDPEIAERMSGLKRRGVDPIAGSGGLDWQ
	1979. .3280	DSGEELAEGLVYAKMSSVDGLANEATLGGFKARNGRNPFPOLIEDFLPTGSETLR
	1979. .3280	LMRYVSAHSGRKALTWKGLRDMAGHSEMSDPOVAADQCCGAAVALPFDHMRQIR
	1979. .3280	TAGAFLPDLDELHSGRSVYAMLKRRIHIEIPLVPMSIST"
BASE COUNT	693 a 967 c 1066 g 814 t	
ORIGIN		
Query Match	8.3%; Score 94.4; DB 1; Length 3540;	
Best Local Similarity	56.2%; Pred. No. 1.8e-13;	
Matches 198; Conservative	0; Mismatches 151; Indels 3; Gaps 1;	
Qy	181 GGTCCGAAAGTTCTGCAATTCGAGAGCCCTTCCTCCGCGAAAGAGCTGATCCGCC	240
Db	2267 GGTGATGATGGTGCCGGTTATGCGGGTTTGCGACAGTGCAGGAAGTGTTGGCTTGCCT	2226
Qy	241 TGTCTGTCGGGAAAAGTCGTGTCACATCTGTGCACACGAAATTTCTCAAGTTGTGTCAT	300
Db	2327 GTCTGTGCGGCAAGATTTCCGCGCACCCGTGTATGACGTGGCCGCTGTGTCCAGATT	2386
Qy	301 CAATCCGGGACATGGAT--CTGTTGCGATGGTGCACATGACCATGCGGCATACAGCTGT	357
Db	2387 GCGGTTGACTCGGCTTCAAGCTGTCAATGCTGACGCTTATCAACGTCAATCATGTCTGT	2446
Qy	358 CAGCGAGCTCCACGACCTATGACCTGGAATTTCCGACAGCTTGAAAGCTGCGACCAACGCT	417
Db	2447 CAGATCTCGCGCGAGCTGTGGGCGTCCGATCGAGTCGGATTGGATGCTGTCAAGAGTGT	2506
Qy	418 CGTCTGTGCGCTACGAAAGCTGAAATGTAACGCTGTGCAACGATACGTCGCGCTGTGAA	477
Db	2507 CGACGGTGCGCAGGAATTTTGGCTCAGCTCGGCTCCAGGGATGGGTCAAGGCGATTGAA	2566
Qy	478 ATCTCTACCGGAAAAACGCGCTGCGACGTCACGCTTACCGGCTACATCAGT	529
Db	2567 GTCAACCCATGGGTGCGCATGGGTGGACGTTACGTGACCGTCTGTCACTCT	2618

LOCUS	APU83788	2439 bp	DNA	circular BCT 15-JUN-1998
DEFINITION	Actinomyces pyogenes plasmid pAP1, complete plasmid sequence.			
ACCESSION	U83788			
VERSION	U83788.1	GI:1805288		
KEYWORDS				
SOURCE				
ORGANISM	Arcanobacterium pyogenes			
	Arcanobacterium pyogenes			
	Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;			
	Actinomycineae; Actinomycetaceae; Arcanobacterium.			
REFERENCE	1 (bases 1 to 2439)			
AUTHORS	Ballington, S.J., Jost, B.H. and Songer, J.G.			
TITLE	The Arcanobacterium (Actinomyces) pyogenes plasmid pAP1 is a member			
JOURNAL	of the p101/pv1 family of rolling circle replication plasmids			
MEDLINE	J. Bacteriol. 180 (12), 3233-3236 (1998)			
PUBMED	9620977			
REFERENCE	2 (bases 1 to 2439)			
AUTHORS	Ballington, S.J., Jost, B.H. and Songer, J.G.			
TITLE	Direct Submission			
JOURNAL	Submitted (06-JUN-1997) Veterinary Science, University of Arizona,			

```

FEATURES
1117 East Lowell Street, Tucson, AZ 85721, USA
source
location/qualifiers
1..2439
/organism="Arcanobacterium pyogenes"
/mol_type="genomic DNA"
/strain="BBR1"
/db_xref="taxon:1661"
/plasmid="pAPI"
1..110
/note="single stranded origin; putative"
39..44
/note="site of single strand initiation; putative"
221..1600
/feature="rep"
221..1600
/feature="rep"
221..1600
/feature="rep"
/feature="replication protein"
/codon_start=1
/transl_table=1
/product="Rep"
/protein_id="AAC46399.1"
/db_xref="GI:1805289"
/translacion="MNNLSERTALSLPAPOLIOKVIIPAAGGRSLKSPFGMTATYARSAGRQ
LASPTESVRKCGRVFAPLVSLDRAISDRGAGYGLHTCGSYMACPVCSAKTAARR
KTDLDQVDAVAKHGMVSMLLTLTORHHKQQLKMLDALSTANRVTSGRPMTEFEK
QFGVLVYRANETTHGKGWVHSHVLIISEKPLSTFVYOROGRRRLPYPEIYIM
SBDIARMEAGLAKHGVDLSDSGGIDMTVAADARAIGNVSMOTSTDAIASEVTLL
GFEKAKNGRTPFOILLADILSLGDVVDLKLMEYKASFGRRALITWSKGLRMANIG
VSGEDDEIASEIEGDEAIALFTHDAMNQVRRFGAELLIDVTSESGRAAAYKMLDFREI
DMSLPIKE"
1673..1676
1685..2023
/note="ORF12"
/codon_start=1
/transl_table=1
/product="unknown"
/protein_id="AAC46400.1"
/db_xref="GI:1805290"
/translacion="MFFIGKVRTFOPIAPCKPSPIDGNGKRTGERSSEKRTQALITG
EDGVELTSLPQFOLLIDAVKRELBEEKYNLAVLRNGGISTYLLTLEAHNV
PAPVVEVSE"
2010..2013
2020..2409
/note="ORF129"
/codon_start=1
/transl_table=1
/product="unknown"
/protein_id="AAC46401.1"
/db_xref="GI:1805291"
/translacion="IMSEPYGISNAELQVFSRPAKREARLYMSRILKELSEGKHS
TPRTALESCEKLYMSVEPRLVDVLEVBGRCRAGNTNSAARSLSPAFIPAFVYOS
LENGRRPAOLLTDDQVAKRKLITGR"
2378..2406
/note="double stranded origin; putative"
BASE COUNT 603 a 657 c 656 g 523 t
ORIGIN
Query Match 7.8% Score 88.8; DB 1; Length 2439;
Blast Local Similarity 55.5% Pred. No. 5.1e-12;
Matches 193; Conservative 0; Mismatches 152; Indels 3; Gaps 14
187 AAAGTTTGATTCGAGAGCGCTTCGTTCTCGCGAAGAGGCTGGATCGCCCTGCTGT 246
|||||
617 AAAGCCCGCATATGATGTTGTCACACTTGTGAAAGCGTCTGGGCGTCCCACTCTGT 676
|||||
247 GCGGAAAAAGTCGGTGCACATCGTCAGACAGAAAATTTCTCAAGTTGTTGTCATCACTC 306
|||||
677 AGCGGAAAATTCGCGCGCTCGCGAAAAACGACCTCCAAAGGTCGTTGACACAGCCGTA 736
|||||
307 GGGACTGA--TCTGTTGCATATGTGACCATGACATGCAATGCCCATACAGCTGTGTACGGG 363
|||||
737 AAACCGGAATAGACCGTCTCAATGCTTAAGCTCACCGACGATCCACCAAGGACATAGGG 796
|||||

```

QY 364 CTCACGACCTATGAGTGAAGCTTTGGAGAGCCTGGAAAGCTGCGAACGCTGCTGCT 423
DB 797 CTAAACACCTCTGGAGAGCCTTGTGCGAGGATGAATCGGCTTACTCTGGTCTGCT 856
QY 424 TGGCGTACGGAAGTGAATGTACGCTGCGAGGATAGTGGCGCTGTGAATCACT 483
DB 857 TGGATTGAGTTCAGGAGGCAATTTGGTTTATGTCGTTATGTTCCACCATGAATTAAT 916
QY 484 CACGGAAGAAAGGCTGGACGCTGACGCTTACGCGCTCTCATGTTTC 531
DB 917 CATGGAAGACGCGCTGGATGTCATTCCTCATTTCTTCTTTC 964

RESULT 9
AY172684/4603 bp DNA circular BCT 12-JAN-2003
LOCUS AY172684
DEFINITION Corynebacterium glutamicum plasmid pAG3, complete sequence.
ACCESSION AY172684
VERSION AY172684.1 GI:27657761
KEYWORDS
SOURCE Corynebacterium glutamicum
ORGANISM Corynebacterium glutamicum
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Corynebacterineae; Corynebacteriaceae; Corynebacterium.
REFERENCE 1 (bases 1 to 4603)
AUTHORS Tauch, A., Puhler, A., Kalinowski, J., and Thierbach, G.
TITLE plasmids in Corynebacterium glutamicum and their molecular
classification by comparative genomics
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 4603)
AUTHORS Tauch, A., Puhler, A., Kalinowski, J., and Thierbach, G.
TITLE Direct Submission
JOURNAL Submitted (05-NOV-2002) Department of Genetics, University of
Bielefeld, Universitaetsstrasse 25, Bielefeld D-33615, Germany
FEATURES
source
1. 4603
/organism="Corynebacterium glutamicum"
/mol_type="genomic DNA"
/strain="22220"
/db_xref="taxon:11718"
/plasmid="pAG3"
complement (200..397)
/note="orf1"
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="hypothetical protein"
/protein_id="AAO18193.1"
/db_xref="GI:27657762"
/translation="MSKOKLSPRAKVLAVDREVTVPVQVAKHGKSGTRQKASALMAL
HKANLIDRTPRGYAPLEVPV"
complement (613..2061)
/gene="repa"
complement (613..2061)
/gene="repa"
/note="repa"
/note="replication protein; essential for pAG3
replication"
/codon_start=1
/transl_table=11
/product="Repa"
/protein_id="AAO18194.1"
/db_xref="GI:27657763"
/translation="MSLSSKQAEHATSAPOATPATHTTGLGTTDTTKHKYKALS
PALFRATLMDLQRTMKTVTEDKTLAGGRVAVGASAVSLVMQAGRAAPWGGI.ONHS
VWSPVASVTCISRRANVDPAIQAMNAGGLTPAHTRGVSTPSDRKQGVSNPV
ERGISLMTLTLRHSKQSLTEVMDALAGQAVNTNAGVSTADKSRVGIAMHY
RALEVTGKGMVHLLHVLPHDRVLSVDRSDSLADVPFMAKARLQMRAPSRB
GIVVAVVAAAGSDAKSIGYTCCKMLSGIAETTTGGVTKKAKGDNTPFOIIGDLGR
QYTRDHALMLEWKGSKGRSGRTGMSGTRDVGINLSDPIDDSIGNDQSEVAVM
VGRGAWGALASDTEKRAVILLDAVAATDSAGARKAADVLKLPVATHTTAIGLAKTD
LDTVLPTHTDVFRELTAEAVWSGRKQTALEVL"
2396..2785

CDS

QY 445 TACGCGTGAGAGGATACGTCGCGCTGTGGAATCACTACGAGAAAACGCTGGCAC 504
DB 1428 TATGGCATTGCCATTCATGACCGGCTATTGAGTCAACCCACGCAAGATGCTGGCAT 1369
QY 505 GTCCAGGTTACGCGCTACTCATGTTCAGTGGTGAAGTGAAGAAATCCCGAATCC 564
DB 1368 GTTCACCTGCATTTGTGTGTTCCATGATCGGAGTGTAAATGTAAGAAAGATTCT 1309
QY 565 TTCTCGATGCGATGTTTCATGTCGATGGAATCCAAATCTGATCTCGGATTTGCTGCG 624
DB 1308 TTGGCTGATGATGTTGTCATAGAGTGCGGCTGAAGGCTGTTGCTGTTGCTATGCGTCA 1249
QY 625 CCCTCTGATTCGCGGCTGCTCATGATAC 655
DB 1248 CCCTCTGATGATGCGGAATGATGTTGTC 1218

RESULT 10
E17316

BASE COUNT 1077 a 1276 c 1112 g 1138 t

ORIGIN

Query Match 4.9%; Score 55.8; DB 1; Length 4603;
Best Local Similarity 54.0%; Pred. No. 0.002;
Matches 114; Conservative 0; Mismatches 97; Indels 0; Gaps 0;

gene
CDS . 4519
complement (4112..4519)
/gene="orf6"
/gene="orf6"
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="hypothetical protein"
/protein_id="AAO18198.1"
/db_xref="GI:27657767"
/translation="MGRNATAMLVSTFQIDNISXALNIDMGSHVAYANASABOL
OSAKLEADKPKRIASLEDSVLTDTCTPGRBROQIMIAISSDKFLKLTSTGAPQ
EVAQADMAALDRFMYATKAHVSPPQPNH"
TAAETVYRAS"
complement (4112..4519)
/gene="orf6"
/gene="orf6"
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="hypothetical protein"
/protein_id="AAO18197.1"
/db_xref="GI:27657766"
/translation="MATPSKPRGLAROLKVRSLSEVYASNVNADHLSDDGKRNA
WRTYAPRHTYIAQYVALETSMSIDAFNAPRKALPTATDTSKIYVELEORI
NNGIPESVDGIKLVTSHPSPRTALIELBARGLSDENVSGILKETSPEVAATE
NMIOHARINSVFGVNLRTMYKALDDBRAIFNMHWVTSADADYDMEVPVTFVPEFKP
TAAETVYRAS"
complement (3344..4030)
/note="orf5"
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="hypothetical protein"
/protein_id="AAO18196.1"
/db_xref="GI:27657765"
/translation="MTKFNRSANDIAAMIGLVIANBNYCTDKOAKTKATAIA
LNKWADELRIRGKDLAHQDETHILERVLRHMLFEFOKATHDNQDLDEHSLQSVHR
IEAGEFTRYGHNWGEVQLAEFDVE"
complement (2840..3322)
/note="orf4"
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="hypothetical protein"
/protein_id="AAO18195.1"
/db_xref="GI:27657764"
/translation="MYRSATLALHAINVRCFVNVNRETLDMTATNIRIKMAT
KAELEQLDKMOALKEQQQLAFNAREBELBOORNLKPVVTVVRVPOYVRDBMKRLAAI
QGMTLQCFANAFAGMIDELSEDOEH"
complement (2840..3322)
/note="orf4"

REFERENCE	1	(bases 1 to 5750)
AUTHORS	O'Riordan, K. and Fitzgerald, G.F.	
TITLE	Molecular characterisation of a 5.75-kb cryptic plasmid from <i>Bifidobacterium breve</i> NCFB 2258 and determination of mode of replication	
JOURNAL	FEMS Microbiol. Lett. 174 (2), 285-294 (1999)	
MEDLINE	99271179	
PUBMED	10339821	
AUTHORS	O'Riordan, K. and Fitzgerald, G.F.	
TITLE	Direct Submision Submitted (22-AUG-1998) Microbiology, University College Cork, Western Road, Cork, Ireland	
JOURNAL	Location/Qualifiers	
FEATURES	1. 5750	
source	/organism="Bifidobacterium breve"	
gene	/mol_type="genomic DNA"	
CDS	/strain="NCFB 2258"	
gene	/db_xref="taxon:1685"	
CDS	/plasmid="pCIBb1"	
gene	1463. .2062	
CDS	/gene="par"	
gene	/note="Par protein; putative involvement in plasmid partition"	
CDS	/codon_start=1	
gene	/transl_table=11	
CDS	/product="partitioning protein"	
gene	/protein_id="AAD34708.1"	
CDS	/db_xref="GI:4972587"	
gene	/translation="MTLSIAKSGGKTSIFLACAAHTADPSLSTILDADQGS	
CDS	ATDMAEABQNDPLPDMKTVAGESDIEIKNEGLTITDTPGFTDLTOTIAKSDL	
gene	CIITTSFQSIDRAKWKTFEALQKAGTSALITRANTKDFRMIDDLANDDANNG	
CDS	ILFDIVRISQSTRYSKPEKTPKRYADADMOEIOKAG"	
gene	complement(2373. .3515)	
CDS	/gene="rep"	
gene	/note="replication gene"	
CDS	complement(2373. .3515)	
gene	/gene="rep"	
CDS	/note="Rep; similar to replication proteins of streptomycetes plasmids and other members of the PC194 family that replicate by a rolling circle mechanism"	
gene	/codon_start=1	
CDS	/transl_table=11	
gene	/product="replication protein"	
CDS	/protein_id="AAD34709.1"	
gene	/db_xref="GI:4972588"	
CDS	/translation="MTNGEKARFTGMLCGSIWACPSCALIRHRAHVALAIGNHA	
gene	EKRKRAADOMQAEHGGRLPPEIMWSDSRGVITRGTTLRHDRMPPLAMTLDAIKG	
CDS	WTKNGASPMQASERBKIRGVPRALEITTYGNGMHPIHFPMLDGDLDGQREMQ	
gene	QMLLDWKTVKRVAAVYKKGQNPYNVAPNDHGDLDLQFKSGKAGTAAATYIKYI	
CDS	GDGQVTLAQBELARQDINKRMGVSVPFQLDSGCLGSLDFQEDLMETWATLRR	
gene	CIITWSGGLKEDEVELEIDEBELAEKADBLPGVGVVPMRVYKDIRKSAPELTALAD	
CDS	AAEREDMOEVARLPGGVILITBEOQDAIADGAKPGDYLPWVSVMV"	
gene	complement(3926. .3934)	
CDS	/note="region conforming to double stranded origin; nicking site of rolling circle plasmids"	
gene	complement(3985. .4845)	
CDS	/gene="ftsK gene homolog"	
gene	complement(3985. .4845)	
CDS	/gene="ftsK gene homolog"	
gene	/note="hypothetical protein; similar to transfer proteins and ftsK homologs"	
CDS	/codon_start=1	
gene	/transl_table=11	
CDS	/product="unknown"	
gene	/protein_id="AAD34710.1"	
CDS	/db_xref="GI:4972589"	
gene	/translation="MKPMWNTKPEPFLRPDPATGITVGKLANGDATLIVT	
CDS	NTSGVAVGGIIPSGKTAQNMVILVALYLSGCCNIVIDKGGDDMGWSEHATTVRG	

QY	371	ACCTATGAGCTGACTTTTCGGCAGCCCTGAAAGCTCGAACAACGGTGTGTTGGCGTA	430
Db	974	CTGTTTGGAATGGATTGCCTTCGGGGTGGCGGAGAATCACTTCGGGGCAAGGGCTGGACGT	915
QY	431	CGGAACGTGAATAATGTCACGGCTGGACGGAATACGTGGCCGCTTTGAAATCATCTCACGGA	490
Db	914	CTAACACAGCTCCGTACACGGGGGTGGAGGGCTTCGTGGTGTGTGAGTTACGACACGGCC	855
QY	491	AAAACGGCTGGACGATCCACGTTCCACGGCGCTACTCATGTTCAATGGTAGCGTAGATA	548
Db	854	GCAATGGATGGCATGTGCAATCTGCATGTGCTCGTGTCTTGTCGGACGACTTCGGCGA	797
RESULT 14			
LOCUS	AF429315	125020 bp	DNA linear PRI 18-JAN-2002
DEFINITION	Homo sapiens junctophilin 3 (JPH3) gene, partial cds.		
ACCESSION	AF429315		
VERSION	AF429315.1	GI:17646244	
KEYWORDS			
SOURCE			
ORGANISM	Homo sapiens (human)		
REFERENCE	Homo sapiens		
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
TITLE	(bases 1 to 125020)		
JOURNAL	Holmes,S.E., O'Hearn,E., Rosenblatt,A., Callahan,C., Hwang,H.S.,		
MEDLINE	Potter,N.T., Ross,C.A. and Margolis,R.L. junctophilin-3 is		
PUBMED	A repeat expansion in the gene encoding junctophilin-3 is associated with Huntington disease--like 2		
REFERENCES	Nat. Genet. 29 (4), 377-378 (2001)		
AUTHORS	2 (bases 1 to 125020)		
TITLE	Holmes,S.E., Jungsoll-Ashworth,R.G., Ross,C.A. and Margolis,R.L.		
JOURNAL	Direct Submission		
MEDLINE	Submitted (05-OCT-2001) Psychiatry, Johns Hopkins Medical		
PUBMED	Institutions, 600 N. Wolfe St., Baltimore, MD 21287, USA		
FEATURES	location/Qualifiers		
source	1..125020		
	/organism="Homo sapiens"		
	/mol_type="genomic DNA"		
	/db_xref="taxon:9606"		
	/chromosome="16"		
	/map_16q24.3; between D16S520 and WI-12410"		
	/note="isolated from a patient with Huntington's Disease-like 2 (HDL2)"		
	complement(35581..35746)		
	/rpt_type=tandem		
	/rpt_unit=cig		
	complement(<36507..>36887)		
	/gene="JPH3"		
	/note="Synonym: JP3"		
	complement(<36507..>36887)		
	/gene="JPH3"		
	/product="Junctophilin 3"		
	complement(<36507..36887)		
	/gene="JPH3"		
	/note="component of the junctional complex between plasma membrane and endoplasmic reticulum"		
	/codon_start=1		
	/product="Junctophilin 3"		
	/protein_id="AA140941.1"		
	/db_xref="GI:17646245"		
	/translation="MSGGKFNFDPDGSGYCGWDEKAHGVCVTPKGQGEYTSWS		
	HGVEVGVTVPSSGNTYQTMAQGKRKHGIGLEBSKRWVVYKGEWTHTPKRGYVRECAQ		
	NGAKYEGTWGLDGDGTEYEDG"		
CDS			
mRNA			
gene			
repeat_region			
BASE COUNT	29056 a 32731 c 30696 g 28283 t 4254 others		
ORIGIN			
Query Match	4.1%; Score 46.4; DB 9; Length 125020;		
Best Local Similarity	10.8%; Pired. No. 0.52;		

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 8, 2003, 12:01:45 ; Search time 281 Seconds
(without alignments)
10951.457 Million cell updates/sec

Title: US-10-007-527A-1

Sequence: 1 atgacacagcgttaagtctga.....taatacgttcgaagtga 1140

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

N_Geneseq.19Jun03.*
1: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.*
2: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.*
3: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.*
4: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.*
5: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.*
6: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.*
7: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.*
8: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.*
9: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.*
10: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.*
11: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.*
12: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.*
13: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.*
14: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.*
15: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.*
16: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.*
17: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.*
18: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.*
19: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.*
20: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.*
21: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.*
22: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.*
23: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.*
24: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.*
25: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	1140	100.0	24	ABQ76122
2	1140	100.0	24	ABQ76124
3	1140	100.0	24	Rhodococcus AN12 d
4	1140	100.0	24	ABQ76126
5	51.6	2297	19	AAV58945
6	37.4	4447	21	AAA90934
7	37.4	4447	21	AAA90951
8	37.4	4447	24	ABU49733

9	37.4	3.3	4447	24	ABU49734	Brevibacterium lac
10	37.4	3.3	4447	24	ABU22582	B. lactofermentum
11	37.4	3.3	4447	24	ABU22583	B. lactofermentum
12	34.8	3.1	1631	21	AA665016	E. coli proliferat
13	34.8	3.1	4590	22	AA424065	Yeast AOP9604-asso
14	34.6	3.0	1505	15	AA055750	Genomic clone G11F
15	34.6	3.0	1954	24	AB199602	Mouse ischaemic co
16	34.4	3.0	1173	22	AA426431	Pseudomonas sp hea
17	34.2	3.0	684	22	AA426442	T. versicolor pyrF
18	34.2	3.0	1314	24	ABQ90222	M. capsulatus gene
19	34.2	3.0	3448	22	AA426441	T. versicolor pyrF
20	34.2	3.0	44938	23	AA595525	Propionibacterium
21	34	3.0	534	24	ABQ41452	Oligonucleotide fo
22	34	3.0	534	24	ABQ41453	Oligonucleotide fo
23	33.6	2.9	786	25	ACA00393	C. glutamicum deri
24	33.6	2.9	832	24	ABQ20710	Oligonucleotide fo
25	33.6	2.9	832	24	ABQ20711	Oligonucleotide fo
26	33.6	2.9	1134	21	AA558791	DNA encoding the b
27	33.6	2.9	1134	24	AB553140	Degenerate DNA enc
28	33.6	2.9	1134	24	AA66165	C glutamicum codin
29	33.6	2.9	1338	22	AA66165	C glutamicum codin
30	33.6	2.9	349980	22	AA688527	DNA encoding novel
31	33.4	2.9	2730	23	AA887782	DNA encoding novel
32	33.4	2.9	3297	23	AA591437	DNA encoding novel
33	33.4	2.9	58985	25	AB259738	Human secreted pro
34	33.4	2.9	143601	25	AB22654	Human epididymal g
35	33.2	2.9	549	21	AA05338	Cat flea head and
36	33.2	2.9	1780	19	AAV52605	Human oxidoreducta
37	33.2	2.9	1783	25	ACC46214	Human dithp intrac
38	33.2	2.9	1798	24	AB156822	Mouse interleukin-
39	33.2	2.9	1843	21	AA337046	Human PRO1265 (UNQ
40	33.2	2.9	1844	20	AA87257	CDNA clone encodin
41	33.2	2.9	1844	21	AA46906	CDNA encoding nove
42	33.2	2.9	1844	22	AA521481	Human CDNA sequenc
43	33.2	2.9	1844	22	AA029290	Human PRO1265 CDNA
44	33.2	2.9	1844	22	AA54260	DNA encoding prote
45	33.2	2.9	1844	22	AA091564	Human PRO1265 CDNA

ALIGNMENTS

RESULT 1
ABQ76122 standard; DNA; 1140 BP.
ID ABQ76122:
AC ABQ76122:
XX
DT 13-JAN-2003 (first entry)
XX
DE Rhodococcus AN12 replication protein Rep DNA.
XX
KW Plasmid stability protein; replication protein; ethylene forming enzyme;
KW carotenoid biosynthesis enzyme; isoprenoid; pyruvate decarboxylase;
KW polyhydroxyalkanoic acid synthase; PBA synthase; nitrile hydratase;
KW alcohol dehydrogenase; terpene synthase; cholesterol oxidase; gene;
KW shuttle vector; Rep; ds.
XX
OS Rhodococcus erythropolis.
XX
PN 200205709-A2.
XX
PD 18-JUL-2002.
XX
PF 12-DEC-2001; 2001WO-US47668.
XX
PR 12-DEC-2000; 2000US-254868P.
XX
PA (DUPO) DU PONT DE NEMOURS & CO E I.
XX
PI Bismucci MG, Cheng Q, Kostichka KN, Tomb J;
DR WPI; 2002-557827/59.

DR P-PSDB; ABB#4278.
XX
PT New nucleic acid molecule encoding replication protein/plasmid
PT stability protein, useful in cloning and expression vectors,
PT particularly shuttle vectors for expression of heterologous genes in
PT Rhodococcus species
XX
PS Claim 2; Page 63-64; 96pp; English.
XX
CC This invention describes a novel nucleic acid encoding a replication
CC protein or a plasmid stability protein. The product of the invention is
CC useful for expression of nucleic acid such as genes encoding enzymes
CC involved in the production of isoprenoid molecules, polyhydroxyalkanoic
CC acid (PHA) synthases, carotenoid biosynthesis enzymes, nitrile
CC hydratases, ethylene forming enzyme, pyruvate decarboxylase, alcohol
CC dehydrogenase, terpene synthases, and cholesterol oxidase in an
CC actinomycetale bacteria. The replication protein or plasmid stability
CC protein are useful in cloning and expression vectors and particularly in
CC shuttle vectors for the expression of homologous and heterologous genes
CC in Rhodococcus sp. This sequence encodes the Rhodococcus AN12 Rep
CC protein described in the disclosure of the invention.
XX
SQ Sequence 1140 BP; 241 A; 287 C; 347 G; 265 T; 0 other;
Query Match 100.0%; Score 1140; DB 24; Length 1140;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1140; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ATGACCGAGCGTAATGCTGAAACACCTTCCGGCAAAACCGGCTCCCGTCTGCTGCG 60
Db 1 ATGACCGAGCGTAATGCTGAAACACCTTCCGGCAAAACCGGCTCCCGTCTGCTGCG 60
QY 61 TCCGATAGCGCGGCATCCGGCAGAACTGCGAACCCAACTTCAACAATCACCAGCTCA 120
Db 61 TCCGATAGCGCGGCATCCGGCAGAACTGCGAACCCAACTTCAACAATCACCAGCTCA 120
QY 121 GAAACATTTAAACGCTTGCGCCGCGCCGATTTCTGCGCTGAAACGCTGACCATTTGTCAC 180
Db 121 GAAACATTTAAACGCTTGCGCCGCGCCGATTTCTGCGCTGAAACGCTGACCATTTGTCAC 180
QY 181 GGTCCGAAAGGTTTGTGATTCGAGAGGCTTGTCTTCCGCGAAAGGCTGATTCGCCCC 240
Db 181 GGTCCGAAAGGTTTGTGATTCGAGAGGCTTGTCTTCCGCGAAAGGCTGATTCGCCCC 240
QY 241 TGCTGTGCGGGAAGGTCGCTGCAATCGTSCAGAACGAAATTTCTCAAGTGTGCTCAT 300
Db 241 TGCTGTGCGGGAAGGTCGCTGCAATCGTSCAGAACGAAATTTCTCAAGTGTGCTCAT 300
QY 301 CAACTCGGAGCTGATCTGTTGCGATGATGATGACCAATGCGCCATACAGCTGTGAC 360
Db 301 CAACTCGGAGCTGATCTGTTGCGATGATGATGACCAATGCGCCATACAGCTGTGAC 360
QY 361 CGGCTCCAGCACTATGAGACTGACCTTCCGCGAGCTGCGAAAGCTGCGACCAAGCTGCT 420
Db 361 CGGCTCCAGCACTATGAGACTGACCTTCCGCGAGCTGCGAAAGCTGCGACCAAGCTGCT 420
QY 421 CGTGTGCGTACGAGAACTGAAATGTAAGCGCTGCGACGAGATGCGGCGCTGTGAATC 480
Db 421 CGTGTGCGTACGAGAACTGAAATGTAAGCGCTGCGACGAGATGCGGCGCTGTGAATC 480
QY 481 ACTCAGGAAAAAAGCGCTGCGACGCTCAGCTGACGCGCTACTCATGTTCAAGTGTGAC 540
Db 481 ACTCAGGAAAAAAGCGCTGCGACGCTCAGCTGACGCGCTACTCATGTTCAAGTGTGAC 540
QY 541 GTGAGTGAAGACATCTCTCGAATCTTCTCGAGTCCGATGTTGATCGGTGAGCTTCCAAA 600
Db 541 GTGAGTGAAGACATCTCTCGAATCTTCTCGAGTCCGATGTTGATCGGTGAGCTTCCAAA 600
QY 601 CTGCTATCTCTGGAGATTGCTGCGCACTAGTAATTCGCGTGTCTGATGTGCAAG 660
Db 601 CTGCTATCTCTGGAGATTGCTGCGCACTAGTAATTCGCGTGTCTGATGTGCAAG 660
QY 661 ATCGCGGCTGAAGCTGATCAAGTTCGCTGCGTATGACGAAATTTGACCTGCGGCTT 720

Db 661 ATCGCGGCTGAAGCTGATCAAGTTCGCTGCGTATGACGAAATTTGACCTGCGGCTT 720
QY 721 GGTATGAGAGTGTGATGAGCGCAAGGAAAGGTCGATGCGAACCGTGCACCTCGG 780
Db 721 GGTATGAGAGTGTGATGAGCGCAAGGAAAGGTCGATGCGAACCGTGCACCTCGG 780
QY 781 GAAATGCTGTTGATGATGAGTGGGCGGAGTCCACAAGCGTTGGAATCTGCGAGAAATT 840
Db 781 GAAATGCTGTTGATGATGAGTGGGCGGAGTCCACAAGCGTTGGAATCTGCGAGAAATT 840
QY 841 GAGTTTGTTCGATGAGGAGCTGCGCAATCGCGTGTCCGATGATGCGTCCGAGCT 900
Db 841 GAGTTTGTTCGATGAGGAGCTGCGCAATCGCGTGTCCGATGATGCGTCCGAGCT 900
QY 901 GGTCTTGGGCGCAAGTAACTAAGTCTCAGATGCTGATGAGAGAAATTCGCCCCGCTC 960
Db 901 GGTCTTGGGCGCAAGTAACTAAGTCTCAGATGCTGATGAGAGAAATTCGCCCCGCTC 960
QY 961 ATGTTTGCATCATCTCCGCGCGATGATGATGATGATGATGATGATGATGATGATGAT 1020
Db 961 ATGTTTGCATCATCTCCGCGCGATGATGATGATGATGATGATGATGATGATGATGAT 1020
QY 1021 TTGCGCGAGATCTCTCGAATCTGTCGAGCTGCGCGGCACTTGGGAAATCTTCTGATCAC 1080
Db 1021 TTGCGCGAGATCTCTCGAATCTGTCGAGCTGCGCGGCACTTGGGAAATCTTCTGATCAC 1080
QY 1081 TTGATTTATGATTTGCGCGAGCGATGTCGCGGCGCCCGATTAATTCGTTGCGAAGTGA 1140
Db 1081 TTGATTTATGATTTGCGCGAGCGATGTCGCGGCGCCCGATTAATTCGTTGCGAAGTGA 1140

RESULT 2
ABQ76124/c
ID ABQ76124 standard; DNA; 6334 BP.
XX
AC ABQ76124;
XX
DT 13-JAN-2003 (first entry)
XX
DE Rhodococcus AN12 derived plasmid pAN12 DNA.
XX
KW Plasmid stability protein; replication protein; ethylene forming enzyme;
KW carotenoid biosynthesis enzyme; isoprenoid; pyruvate decarboxylase;
KW polyhydroxyalkanoic acid synthase; PHA synthase; nitrile hydratase;
KW alcohol dehydrogenase; terpene synthase; cholesterol oxidase;
KW shuttle vector; circular; ds.
XX
OS Rhodococcus erythropolis.
OS Synthetic.
OS
PN WO200255709-A2.
PD 18-JUL-2002.
XX
PF 12-DEC-2001; 2001WO-US47868.
PR 12-DEC-2000; 2000US-254868P.
XX
PA (DUPO) DU PONT DE NEMOURS & CO E I.
PI Bramucci MG, Cheng Q, Kostichka KN, Tomb J;
XX
DR WPI, 2002-557827/59.
XX
XX New nucleic acid molecule encoding replication protein/plasmid
PT stability protein, useful in cloning and expression vectors,
PT particularly shuttle vectors for expression of heterologous genes in
PT Rhodococcus species
PS Claim 19; Page 68-71; 96pp; English.
CC This invention describes a novel nucleic acid encoding a replication

CC protein or a plasmid stability protein. The product of the invention is
 CC useful for expression of nucleic acid such as genes encoding enzymes
 CC involved in the production of isoprenoid molecules, polyhydroxyalkanoic
 CC acid (PHA) synthases, carotenoid biosynthesis enzymes, nitrile
 CC hydratases, ethylene forming enzyme, pyruvate decarboxylase, alcohol
 CC dehydrogenase, terpene synthases, and cholesterol oxidase in an
 CC Actinomycetales bacteria. The replication protein or plasmid stability
 CC protein are useful in cloning and expression vectors and particularly in
 CC shuttle vectors for the expression of homologous and heterologous genes
 CC in Rhodococcus sp. This sequence represents the Rhodococcus AN12 derived
 CC plasmid PAN12 DNA described in the disclosure of the invention.

XX
 SQ Sequence 6334 BP; 1386 A; 1807 C; 1681 G; 1460 T; 0 other;

Query Match 100.0%; Score 1140; DB 24; Length 6334;
 Best Local Similarity 100.0%; Pred. No. 0;

Matches 1140; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGACCGAGCTAAGTGTGTAACACCTTTCCGCGAAGACCGGCTCCCGTCTGTGTG 60
 DB 3051 ATGACCGAGCTAAGTGTGTAACACCTTTCCGCGAAGACCGGCTCCCGTCTGTGTG 2992
 QY 61 TCCGATTAAGCGCGGCATCCGCGACGAACTGCGAACCCTTCAACAAATCACCGTCA 120
 DB 2991 TCCGATTAAGCGCGGCATCCGCGACGAACTGCGAACCCTTCAACAAATCACCGTCA 2932
 QY 121 GAAACATTTAACCGCTGTGCGCGCGCGATTTCTGCGGTGAACGCTGTGACATTTGCAAC 180
 DB 2931 GAAACATTTAACCGCTGTGCGCGCGCGATTTCTGCGGTGAACGCTGTGACATTTGCAAC 2872
 QY 181 GGTCCGAAGGTTCTGGATTTCGAGGCTTCTGCTGCGGAAAAGGCTGTGATTCGCC 240
 DB 2871 GGTCCGAAGGTTCTGGATTTCGAGGCTTCTGCTGCGGAAAAGGCTGTGATTCGCC 2812
 QY 241 TGCCTGCGCGGAAAAGTCCGTGACATCGTGAAGAGAAATTTCTCAAGTTGTGTCAT 300
 DB 2811 TGCCTGCGCGGAAAAGTCCGTGACATCGTGAAGAGAAATTTCTCAAGTTGTGTCAT 2752
 QY 301 CAACTCGGAGCTGGATCTGTGCGATGCTGACATGACATGCGCCATACAGCTGTGAC 360
 DB 2751 CAACTCGGAGCTGGATCTGTGCGATGCTGACATGACATGCGCCATACAGCTGTGAC 2692
 QY 361 CGGCTCCAGACCTATGATGACTGACTTTCCGCGAGCTTGAAGAGTGTGACCAACGGTGT 420
 DB 2691 CGGCTCCAGACCTATGATGACTGACTTTCCGCGAGCTTGAAGAGTGTGACCAACGGTGT 2632
 QY 421 CGTTGCGCTAGCGAAGCTGAAATGTAAGTGTGCGACGGAATACGTCGCGCTGTGTAATC 480
 DB 2631 CGTTGCGCTAGCGAAGCTGAAATGTAAGTGTGCGACGGAATACGTCGCGCTGTGTAATC 2572
 QY 481 ACTCAGGAAAAAAGCGCTGGACGTCACGCTTCAACGCTTCACTCATGTCAGTGTGAC 540
 DB 2571 ACTCAGGAAAAAAGCGCTGGACGTCACGCTTCAACGCTTCACTCATGTCAGTGTGAC 2512
 QY 541 GTGAGTGAAGAACATCTCTCGAATCTTTCCGATGCGATGTTGATCGGTGACTTCCAA 600
 DB 2511 GTGAGTGAAGAACATCTCTCGAATCTTTCCGATGCGATGTTGATCGGTGACTTCCAA 2452
 QY 601 CTGCTATCTCTGGAGTTTGTGCGCGCACTAAGTAATCGGCTGTGCTGATGTAAGAA 660
 DB 2451 CTGCTATCTCTGGAGTTTGTGCGCGCACTAAGTAATCGGCTGTGCTGATGTAAGAA 2392
 QY 661 ATCGCGGATGAAGCTGATCAAGTTCTGCTGCTGATGTCGAAATTTGATCGGCGTT 720
 DB 2391 ATCGCGGATGAAGCTGATCAAGTTCTGCTGCTGATGTCGAAATTTGATCGGCGTT 2332
 QY 721 GGTATGAGAGTTGGTGTAGTGGCGACGAAAAAGTGTGACATGCAACCGTGCACCTGG 780
 DB 2331 GGTATGAGAGTTGGTGTAGTGGCGACGAAAAAGTGTGACATGCAACCGTGCACCTGG 2272
 QY 781 GAAATGCTGTGATGATGAGTGGCGGGAATCAACAACGTTGGAACGTGTGCGGGAATTT 840
 DB 2271 GAAATGCTGTGATGATGAGTGGCGGGAATCAACAACGTTGGAACGTGTGCGGGAATTT 2212

QY 841 GAGTTTGTTCATGAGGACGTCGGGCAATCGCGTGTCCCGTGAATTCGTCGCCGAGCT 900
 DB 2211 GAGTTTGTTCATGAGGACGTCGGGCAATCGCGTGTCCCGTGAATTCGTCGCCGAGCT 2152
 QY 901 GGTCTGGGAGGAACTAACAGATGCTGATCGTTGAGCAGAGAAATCTGCCCGGTC 960
 DB 2151 GGTCTGGGAGGAACTAACAGATGCTGATCGTTGAGCAGAGAAATCTGCCCGGTC 2092
 QY 961 ATGTTTGCATGATTCGCGCGGATCGTGAATGATGATTCGACCTTGTGCGCTTACATC 1020
 DB 2091 ATGTTTGCATGATTCGCGCGGATCGTGAATGATGATTCGACCTTGTGCGCTTACATC 2032
 QY 1021 TTGCGCGAATCTCTGCAATCTGTCGAGCTGGCGCGCATTTGGAAAAATCTTGATAC 1080
 DB 2031 TTGCGCGAATCTCTGCAATCTGTCGAGCTGGCGCGCATTTGGAAAAATCTTGATAC 1972
 QY 1081 TTGCAATTATGATTCGCGCGAGCGATGAGCGGCGCCGATTAATTCGTTGCAAGTA 1140
 DB 1971 TTGCAATTATGATTCGCGCGAGCGATGAGCGGCGCCGATTAATTCGTTGCAAGTA 1912

RESULT 3

ABQ76126/c
 ID ABQ76126 standard; DNA; 9652 BP.

XX ABQ76126;

DT 13-JAN-2003 (first entry)

DE Plasmid pRHR171 DNA.

XX Plasmid stability protein; replication protein; ethylene forming enzyme;
 XX carotenoid biosynthesis enzyme; isoprenoid; pyruvate decarboxylase;
 XX polyhydroxyalkanoic acid synthase; PHA synthase; nitrile hydratase;
 XX alcohol dehydrogenase; terpene synthase; cholesterol oxidase;
 XX shuttle vector; circular; ds.

XX Synthetic.

XX MO200255709-A2.

PD 18-JUL-2002.

PF 12-DEC-2001; 2001WO-US47868.

PR 12-DEC-2000; 2000US-254868P.

XX (DUPO) DU PONT DE NEMOURS & CO E I.

XX Bramucci MG, Cheng Q, Kostichka KN, Tomb J;

XX WPI; 2002-557827/59.

XX New nucleic acid molecule encoding replication protein/plasmid

PT stability protein, useful in cloning and expression vectors,

PT particularly shuttle vectors for expression of heterologous genes in

PT Rhodococcus species -

PS Claim 27; Page 72; 96pp; English.

XX This invention describes a novel nucleic acid encoding a replication
 CC protein or a plasmid stability protein. The product of the invention is
 CC useful for expression of nucleic acid such as genes encoding enzymes
 CC involved in the production of isoprenoid molecules, polyhydroxyalkanoic
 CC acid (PHA) synthases, carotenoid biosynthesis enzymes, nitrile
 CC hydratases, ethylene forming enzyme, pyruvate decarboxylase, alcohol
 CC dehydrogenase, terpene synthases, and cholesterol oxidase in an
 CC Actinomycetales bacteria. The replication protein or plasmid stability
 CC protein are useful in cloning and expression vectors and particularly in
 CC shuttle vectors for the expression of homologous and heterologous genes
 CC in Rhodococcus sp. This sequence represents the Plasmid pRHR171 DNA
 CC described in the disclosure of the invention.

XX Sequence 9652 BP; 2097 A; 2729 C; 2559 G; 2267 T; 0 other;
SQ
Query Match 100.0%; Score 1140; DB 24; Length 9652;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1140; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGACGAGCGTAAAGTCTGAAACCTTTCCGGGAAAAGCCGCGCTCCGCTCGTGTG 60
DB 5052 ATGACGAGCGTAAAGTCTGAAACCTTTCCGGGAAAAGCCGCGCTCCGCTCGTGTG 4993

QY 61 TCCGATTAAGCGCGGCATCCGCGACGAACCTGACCCAACTTAAACAATCAACAGTCA 120
DB 4992 TCCGATTAAGCGCGGCATCCGCGACGAACCTGACCCAACTTAAACAATCAACAGTCA 4933

QY 121 GAACATTTAAGCCTGTGGCCGCGCATTTCTGGCGTGAACGGTGTGACATTGTCAAC 180
DB 4932 GAACATTTAAGCCTGTGGCCGCGCATTTCTGGCGTGAACGGTGTGACATTGTCAAC 4873

QY 181 GGTCCGAAAAGTCTGGATTGCGAGGCGCTTCTGCTGCGAAAAGGGCTGGATCTGCCCC 240
DB 4872 GGTCCGAAAAGTCTGGATTGCGAGGCGCTTCTGCTGCGAAAAGGGCTGGATCTGCCCC 4813

QY 241 TGTCTGTGCGGAAAAAGTCCGCTGACATCTGCGACGAAGAAATTTCTCAAGTTGTGCTCAT 300
DB 4812 TGTCTGTGCGGAAAAAGTCCGCTGACATCTGCGACGAAGAAATTTCTCAAGTTGTGCTCAT 4753

QY 301 CAACCTGGGAGCTGGATCTGTGGATGATGATGATGATGATGATGATGATGATGATGATGAT 360
DB 4752 CAACCTGGGAGCTGGATCTGTGGATGATGATGATGATGATGATGATGATGATGATGATGAT 4693

QY 361 CGGCTCCACGACCTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 420
DB 4692 CGGCTCCACGACCTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 4633

QY 421 CGTTGGCGTTCGGAACGTGAATGTATGATGATGATGATGATGATGATGATGATGATGATGAT 480
DB 4632 CGTTGGCGTTCGGAACGTGAATGTATGATGATGATGATGATGATGATGATGATGATGATGAT 4573

QY 481 ACTCAGGAAAAAAGCGCTGGACGTCACGTCACGCTGACCTGACCTGACCTGACCTGACCTGAC 540
DB 4572 ACTCAGGAAAAAAGCGCTGGACGTCACGTCACGCTGACCTGACCTGACCTGACCTGACCTGAC 4513

QY 541 GTGATGAGAACATCTCGATCTCTTCGATGCGATGCGATGCGATGCGATGCGATGCGATGCGAT 600
DB 4512 GTGATGAGAACATCTCGATCTCTTCGATGCGATGCGATGCGATGCGATGCGATGCGATGCGAT 4453

QY 601 CTGCTATCTCTGGGATTTGCTGCGCGCATGATGATGATGATGATGATGATGATGATGATGATGAT 660
DB 4452 CTGCTATCTCTGGGATTTGCTGCGCGCATGATGATGATGATGATGATGATGATGATGATGATGAT 4393

QY 661 ATGCGGCGTGAAGCTGATCAAGTTCTGCGCGCATGATGATGATGATGATGATGATGATGATGATGAT 720
DB 4392 ATGCGGCGTGAAGCTGATCAAGTTCTGCGCGCATGATGATGATGATGATGATGATGATGATGATGAT 4333

QY 721 GGTATGAGAGTTGTGATGCGACGCGAAAAAGTGTGACATGCGAACCGTGCACCTGTG 780
DB 4332 GGTATGAGAGTTGTGATGCGACGCGAAAAAGTGTGACATGCGAACCGTGCACCTGTG 4273

QY 781 GAAATCGCTGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 840
DB 4272 GAAATCGCTGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 4213

QY 841 GAGTTGGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 900
DB 4212 GAGTTGGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 4153

QY 901 GGTCTTGGGCGAGAACATCAAGATGCTCAGATCTTGAAGCAGAAAGATCTGCCCCGCTC 960
DB 4152 GGTCTTGGGCGAGAACATCAAGATGCTCAGATCTTGAAGCAGAAAGATCTGCCCCGCTC 4093

QY 961 ATGTTGGCATCATTTCCGCGCGCATGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1020

DB 4092 ATGTTGGCATCATTTCCGCGCGCATGCTGATGATGATGATGATGATGATGATGATGATGATGAT 4033

QY 1021 TTCCGCGAGATCTTCGACCTGTCGACCTGCGCGCATTTGGAAAAATCTTGATGATCAC 1080

DB 4032 TTCCGCGAGATCTTCGACCTGTCGACCTGCGCGCATTTGGAAAAATCTTGATGATCAC 3973

QY 1081 TTGCAATTATTCATTTCCCGCGAGCGGATGTGCGGCCCCCGGATTAATTCGTTTCGCAAGTGA 1140

DB 3972 TTGCAATTATTCATTTCCCGCGAGCGGATGTGCGGCCCCCGGATTAATTCGTTTCGCAAGTGA 3913

RESULT 4

AB076125/c
ID AB076125 standard; DNA; 11241 BP.

AB076125;

13-JAN-2003 (first entry)

Plasmid pRHB17 DNA.

Plasmid stability protein; replication protein; ethylene forming enzyme;

carotenoid biosynthesis enzyme; isoprenoid; pyruvate decarboxylase;

KM polyhydroxyalkanoic acid synthase; pHA synthase; nitrile hydratase;

KM alcohol dehydrogenase; terpene synthase; cholesterol oxidase;

KM shuttle vector; circular; ds.

OS Synthetic.

PN ~~200255709-A2.~~

PD 18-UTL-2002.

PF 12-DEC-2001; 2001MO-US47868.

PR 12-DEC-2000; 2000US-254868P.

PA (DUPO) DU POINT DE MEMOIRS & CO E. I.

PI Bramucci MG, Cheng Q, Kostichka KM, Tomb J;

DR WPI; 2002-557827/59.

XX Claim 26; Page 71-72; 96pp; English.

XX This invention describes a novel nucleic acid encoding a replication

protein or a plasmid stability protein. The product of the invention is

involved in the production of isoprenoid molecules, polyhydroxyalkanoic

acid (PHA) synthases, carotenoid biosynthesis enzymes, nitrile

hydratases, ethylene forming enzyme, pyruvate decarboxylase, alcohol

CC dehydrogenase, terpene synthases, and cholesterol oxidase in an

CC Actinomycetalea bacteria. The replication protein or plasmid stability

CC protein are useful in cloning and expression vectors and particularly in

CC shuttle vectors for the expression of homologous and heterologous genes

CC in Rhodococcus sp. This sequence represents the Plasmid pRHB17 DNA

XX described in the disclosure of the invention.

SQ Sequence 11241 BP; 2517 A; 3109 C; 2945 G; 2670 T; 0 other;

Query Match 100.0%; Score 1140; DB 24; Length 11241;

Best Local Similarity 100.0%; Pred. No. 0; Mismatches 0; Indels 0; Gaps 0;

Matches 1140; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGACGAGCGTAAAGTCTGAAACCTTTCCGGGAAAAGCCGCGCTCCGCTCGTGTG 60

DB 6641 ATGACGAGCGTAAAGTCTGAAACCTTTCCGGGAAAAGCCGCGCTCCGCTCGTGTG 6582

```

QY 61 TCCGATTAAGCGGGCATCCGGGCGAAGTGGGACCCAACTTCAACAAATACACAGTCA 120
DB 6581 TCCGATTAAGCGGGCATCCGGGCGAAGTGGGACCCAACTTCAACAAATACACAGTCA 6522
QY 121 GAAACATTTAAAGCCCTGGGCGCGATTTCTGGGCGTGAACCGGTGACCATTTGGCAAC 180
DB 6521 GAAACATTTAAAGCCCTGGGCGCGATTTCTGGGCGTGAACCGGTGACCATTTGGCAAC 6462
QY 181 GGTCCGAAAAGTTCTGGAATTCGAGAGGCTTGTCTCTGGCGAAAAGGCTGATCTCCCC 240
DB 6461 GGTCCGAAAAGTTCTGGAATTCGAGAGGCTTGTCTCTGGCGAAAAGGCTGATCTCCCC 6402
QY 241 TGGTGGCGGAAAAGTCGGTGACATCGTGACAGCAAAATTTCTCAAGTTGGTCTCAT 300
DB 6401 TGGTGGCGGAAAAGTCGGTGACATCGTGACAGCAAAATTTCTCAAGTTGGTCTCAT 6342
QY 301 CAACCTGGGACTGGATCTGTGGGATGGTGAACATACATGCGCATACAGTGGTCAAG 360
DB 6341 CAACCTGGGACTGGATCTGTGGGATGGTGAACATACATGCGCATACAGTGGTCAAG 6282
QY 361 CGGCTCCAGCACTTATGACTGAGACTTTGGGCGAGCTTGGAAAAGCTGCGAACAGTCCGT 420
DB 6281 CGGCTCCAGCACTTATGACTGAGACTTTGGGCGAGCTTGGAAAAGCTGCGAACAGTCCGT 6222
QY 421 CGTTGGCGTACCGGAACGTGAATGTACGCGTGACAGATACGTGCGCGCTGTGAATC 480
DB 6221 CGTTGGCGTACCGGAACGTGAATGTACGCGTGACAGATACGTGCGCGCTGTGAATC 6162
QY 481 ACTCAAGGAAAAAAGCGGCTGGGCGCGTCCAGCTTCAAGCTTCAATGTCAGTGGTGCAC 540
DB 6161 ACTCAAGGAAAAAAGCGGCTGGGCGCGTCCAGCTTCAAGCTTCAATGTCAGTGGTGCAC 6102
QY 541 GTGAGTGAAGAACATCTCGAATCTTCTCGATGCGATGATGTCAGTGGTGAACCTTCAAA 600
DB 6101 GTGAGTGAAGAACATCTCGAATCTTCTCGATGCGATGATGTCAGTGGTGAACCTTCAAA 6042
QY 601 CTGGAATCTCTGGGAAATTTGGTGGCGCACTACGTAATTCGGGTGCTCGATGTACGAAA 660
DB 6041 CTGGAATCTCTGGGAAATTTGGTGGCGCACTACGTAATTCGGGTGCTCGATGTACGAAA 5982
QY 661 ATGCGGCGTGAAGCTGATCAAGTTCTGCGTGGCTGATCTGACGAAAATTCGATCTGGCGTT 720
DB 5981 ATGCGGCGTGAAGCTGATCAAGTTCTGCGTGGCTGATCTGACGAAAATTCGATCTGGCGTT 5922
QY 721 GGTATGAGAGTTGGTATGGGCGACGCGAAAAGTGTGCAATGCGCAACCGTGCACCTCGG 780
DB 5921 GGTATGAGAGTTGGTATGGGCGACGCGAAAAGTGTGCAATGCGCAACCGTGCACCTCGG 5862
QY 781 GAAATCGCTGTTGATGAGTGGGCGGCGATCCAAAGCTTTGGAACCTGTGGCGAGAAATTT 840
DB 5861 GAAATCGCTGTTGATGAGTGGGCGGCGATCCAAAGCTTTGGAACCTGTGGCGAGAAATTT 5802
QY 841 GAATTTGGTTCGATGGGACGTCGCGGCAATGCGGTGGTCCCGTGGATTTGGCTGCCAGCT 900
DB 5801 GAATTTGGTTCGATGGGACGTCGCGGCAATGCGGTGGTCCCGTGGATTTGGCTGCCAGCT 5742
QY 901 GGTCTTGGGCGCAACATCAAGATGCTCAGATGCTTGAAGAGGAAAGATTTGCCCGGCTC 960
DB 5741 GGTCTTGGGCGCAACATCAAGATGCTCAGATGCTTGAAGAGGAAAGATTTGCCCGGCTC 5682
QY 961 ATGTTGCGATTCATTCGCGCGCATCGTGATGATGATTCGACTTGTGGCGCTTACGTC 1020
DB 5681 ATGTTGCGATTCATTCGCGCGCATCGTGATGATGATTCGACTTGTGGCGCTTACGTC 5622
QY 1021 TTGCGGCGAGATCTCTCGGACTGTGCGAAGCTGCGCGCATTTGGGAAAATCTTGTGATCAC 1080
DB 5621 TTGCGGCGAGATCTCTCGGACTGTGCGAAGCTGCGCGCATTTGGGAAAATCTTGTGATCAC 5562
QY 1081 TTGCAATTATGATTTGCCCGGAGGAGATGCGGCGCCCGGATAATTCGATTCGCAAGTGA 1140
DB 5561 TTGCAATTATGATTTGCCCGGAGGAGATGCGGCGCCCGGATAATTCGATTCGCAAGTGA 5502

```

```

RESULT 5
AAV58945
ID AAV58945 standard; DNA; 2297 BP.
XX
AC AAV58945;
XX
DT 05-JAN-1999 (first entry)
XX
DE B. breve essential region gene.
XX
DE Essential region gene; shuttle vector; ds.
XX
OS Bifidobacterium breve.
XX
FH Key 387..392 Location/Qualifiers
FT -35_signal /tag= a
FT -10_signal 410..415 /tag= b
FT RBS 525..530 /tag= c
FT CDS 539..2023 /tag= d
FT
XX
PN JP10262670-A.
XX
PD 06-OCT-1998.
XX
PF 27-MAR-1997; 97JP-0091387.
XX
PR 27-MAR-1997; 97JP-0091387.
XX
PA (HONS ) YAKULT HONSHA KK.
XX
DR MPI: 1998-587288/50.
XX
DR P-PSDB; AAW73071.
XX
PT Shuttle vector for a Bifidobacterium species - contains specific
PT elements from existing plasmids e.g. pNB1 of ATCC 15698
XX
PS Disclosure; Page 14-15; 17pp; Japanese.
XX
CC This sequence represents the essential region gene from plasmid pNB1
CC used in the vector of the invention. The vector is a shuttle vector for a
CC Bifidobacterium, and contains: (a) a replicated essential region
CC originated from a plasmid pNB1 of Bifidobacterium breve ATCC 15698;
CC (b) a replicated essential region originated from a plasmid of E. coli;
CC and (c) an antibiotic-resistant gene originating from a plasmid of E. coli
CC and an antibiotic-resistant gene functioning by a Bifidobacterium. The
CC shuttle vector can be used to transform various Bifidobacteria species.
XX
SQ Sequence 2297 BP; 524 A; 659 C; 689 G; 425 T; 0 other;
Query Match 4.5%; Score 51.6; DB 19; Length 2297;
Best Local Similarity 52.9%; Pred. No. 1.8e-05;
Matches 111; Conservative 0; Mismatches 99; Indels 0; Gaps 0;
QY 394 GCGTGAAGAGCTGCGACCAAGCTGCTGGGCGGTAACGAAATGTAAGGCTGC 453
DB 1184 GCGTGAAGAGAGTGAATGTAACGAAAGCTTGGCAACGGGCTCGGAAGTTGAAAATC 1243
QY 454 GACGATACGTGGCGCTGTTGAAATCACTCAAGGAAAAACGCTGGACGTCACGTT 513
DB 1244 AGGGGTTTGTCCGCCGATGTAATCACTACGATGTAACGGCTGGACCTTCACATT 1303
QY 514 CACGCGCTACTATGTTCAAGTGTGACGTGAAGTGAACATCTCGAATCTTTCGGAT 573
DB 1304 CATTTCGATGTTTCTCGATGCGCATGACGATGCGGACGATGAGGCAATGACAGCA 1363
QY 574 GCGATGTTGATCGGTGACCTTCCAAATC 603
DB 1364 TGGCTGCTGATCGCTGGAACCAATGGTC 1393

```

```
RESULT 6
AAA90934
ID AAA90934 standard; DNA; 4447 BP.
XX
AC AAA90934;
XX
DT 15-JAN-2001 (first entry)
XX
DE B. lactofermentum p48K coding sequence.
XX
KM Temperature sensitive plasmid; TSRCR; protein production;
XX temperature sensitive replication control region; p48K; ds.
XX
OS Brevibacterium lactofermentum.
XX
FH Key Location/Qualifiers
FT CDS 1318..2601
FT /*tag= a
XX
PN EP1038966-A1.
XX
PD 27-SEP-2000.
XX
PF 16-MAR-2000; 2000EP-0105326.
XX
PR 16-MAR-1999; 99JP-0069896.
XX
PA (AJIN ) AJINOMOTO CO INC.
XX
PI Nakamura J, Kanno S, Kimura E, Matsui K, Nakamatsu T;
XX
DR WPI; 2000-573832/54.
XX
DR P-PSDB; AAY97536.
XX
PT Plasmids containing a temperature sensitive replication control regions
PT useful for breeding microorganisms for the production of amino acids by
PT fermentation -
XX
PS Claim 7; Page 14-17; 29pp; English.
XX
XX This sequence encodes the Brevibacterium lactofermentum p48K protein.
XX The invention relates to a plasmid containing a temperature sensitive
XX replication control region (TSRCR) and a marker gene (MG). The TSRCR is
XX derived from plasmid pAM330 harboured by Brevibacterium lactofermentum
XX (ATCC 13869) and allows the plasmid to replicate autonomously at a low
XX temperature but does not allow the plasmid to replicate autonomously at
XX an elevated temperature in coryneform bacteria within a temperature
XX range in which the bacteria can grow. The plasmid can be used for
XX modifying a chromosomal gene in a coryneform bacterium, which may be used
XX for the production of useful substances, such as amino acids, by
XX fermentation, to change their genetic traits. Therefore, the plasmid can
XX be used for breeding microorganisms for the production of amino acids by
XX fermentation. The plasmid comprises a TSRCR that allows the plasmid to
XX replicate autonomously at an elevated temperature in coryneform
XX bacteria within a temperature range in which the bacteria can grow.
XX
SQ Sequence 4447 BP; 824 A; 997 C; 1378 G; 1248 T; 0 other;
XX
Query Match 3.3%; Score 37.4; DB 21; Length 4447;
Best Local Similarity 53.0%; Pred. No. 0.87;
Matches 80; Conservative 0; Mismatches 71; Indels 0; Gaps 0;
```

```
DB 1915 GGTATGACGCCCATGCTGAGACAGGGG 1945
RESULT 7
AAA90951
ID AAA90951 standard; DNA; 4447 BP.
XX
AC AAA90951;
XX
DT 15-JAN-2001 (first entry)
XX
DE B. lactofermentum pSFK6 coding sequence.
XX
KM Temperature sensitive plasmid; TSRCR; protein production;
XX temperature sensitive replication control region; pSFK6; ds.
XX
OS Brevibacterium lactofermentum.
XX
FH Key Location/Qualifiers
FT CDS 1318..2601
FT /*tag= a
XX
PN EP1038966-A1.
XX
PD 27-SEP-2000.
XX
PF 16-MAR-2000; 2000EP-0105326.
XX
PR 16-MAR-1999; 99JP-0069896.
XX
PA (AJIN ) AJINOMOTO CO INC.
XX
PI Nakamura J, Kanno S, Kimura E, Matsui K, Nakamatsu T;
XX
DR WPI; 2000-573832/54.
XX
DR P-PSDB; AAY97537.
XX
PT Plasmids containing a temperature sensitive replication control regions
PT useful for breeding microorganisms for the production of amino acids by
PT fermentation -
XX
PS Disclosure; Page 19-22; 29pp; English.
XX
XX This sequence encodes the Brevibacterium lactofermentum pSFK6 protein.
XX The invention relates to a plasmid containing a temperature sensitive
XX replication control region (TSRCR) and a marker gene (MG). The TSRCR is
XX derived from plasmid pAM330 harboured by Brevibacterium lactofermentum
XX (ATCC 13869) and allows the plasmid to replicate autonomously at a low
XX temperature but does not allow the plasmid to replicate autonomously at
XX an elevated temperature in coryneform bacteria within a temperature
XX range in which the bacteria can grow. The plasmid can be used for
XX modifying a chromosomal gene in a coryneform bacterium, which may be used
XX for the production of useful substances, such as amino acids, by
XX fermentation, to change their genetic traits. Therefore, the plasmid can
XX be used for breeding microorganisms for the production of amino acids by
XX fermentation. The plasmid comprises a TSRCR that allows the plasmid to
XX replicate autonomously at an elevated temperature in coryneform
XX bacteria within a temperature range in which the bacteria can grow.
XX
SQ Sequence 4447 BP; 827 A; 993 C; 1375 G; 1252 T; 0 other;
XX
Query Match 3.3%; Score 37.4; DB 21; Length 4447;
Best Local Similarity 53.0%; Pred. No. 0.87;
Matches 80; Conservative 0; Mismatches 71; Indels 0; Gaps 0;
```

QY 613 GGATTTCGCCCACTAGCTATTCGGG 643
DB 1915 GGTATGACGCGCCACTGCGTGACGCGG 1945

RESULT 8

ID ABL49733

ABL49733 standard; DNA; 4447 BP.

AC ABL49733;

DT 29-MAY-2002 (first entry)

DE Brevibacterium lactofermentum pSFK6 nucleotide sequence SEQ ID NO:5.

KM Kanamycin resistant gene; mutation; high temperature resistance;

KM coryneform microbe; transformation; gene; ds.

OS Brevibacterium lactofermentum.

FH Key Location/Qualifiers

FT CDS 1318..2601

FT /*tag= a

FT /product= "pSFK6"

PN JP2002017362-A.

PD 22-JAN-2002.

PF 05-JUL-2000; 2000JP-0204236.

PR 05-JUL-2000; 2000JP-0204236.

PA (AJIN) AJINOMOTO KK.

DR MPI; 2002-263243/31.

DR P-PSDB; ABB06341.

PT Transformation of high temperature-resistant coryneform microbe and

PT transformant -

PS Example; Page 16-18; 28pp; Japanese.

CC The present invention describes a method for transforming a high

CC temperature-resistant coryneform microbe in which the high temperature-

CC resistant coryneform microbe is treated with a chemical changing the

CC structure of cell wall of the microbe and an electric pulse is applied

CC to a solution containing the microbe having a changed structure of cell

CC wall and a DNA. Also described are: (1) transforming a high temperature-

CC temperature-resistant coryneform microbe in which the cell wall of a high

CC is applied to a solution containing the microbe having weakened cell

CC wall and a DNA; (2) a transformant of a high temperature-resistant

CC coryneform microbe in which a recombinant DNA is introduced; and

CC (3) a transformant of a high temperature-resistant coryneform microbe

CC carrying both a plasmid vector derived from the high temperature-

CC resistant coryneform microbe and a plasmid vector derived from

CC Corynebacterium glutamicum or Brevibacterium lactofermentum. The method

CC is used for transforming a high temperature-resistant coryneform microbe.

CC The present sequence encodes a Brevibacterium lactofermentum pSFK6

CC protein sequence, which is used in an example from the present

CC invention.

SQ Sequence 4447 BP; 824 A; 997 C; 1378 G; 1248 T; 0 other;

Query Match 3.3%; Score 37.4; DB 24; Length 4447;

Best Local Similarity 53.0%; Pred. No. 0.87; Indels 0; Gaps 0;

Matches 80; Conservative 0; Mismatches 71; Indels 0; Gaps 0;

QY 493 AACGGCTGACGCGCCACTAGCTATTCGGG 552
DB 1795 AACGGCTGACGCGCCACTAGCTATTCGGG 1054

QY 553 ATCCGAAATCTTCTCGAGATGCTTCAATCGGTGACCTTCAACTGCTATCTCG 612
DB 1855 GAATCAAGGCGCTTTGAGGATTCAGTTTTCCCGCTGCTGCTGTGTTAAGGCC 1914

QY 613 GGATTTCGCCCACTAGCTATTCGGG 643
DB 1915 GGTATGACGCGCCACTGCGTGACGCGG 1945

RESULT 9

ID ABL49734

ABL49734 standard; DNA; 4447 BP.

AC ABL49734;

DT 29-MAY-2002 (first entry)

DE Brevibacterium lactofermentum p48K nucleotide sequence SEQ ID NO:7.

KM Kanamycin resistant gene; mutation; high temperature resistance;

KM coryneform microbe; transformation; gene; ds.

OS Brevibacterium lactofermentum.

FH Key Location/Qualifiers

FT CDS 1318..2601

FT /*tag= a

FT /product= "p48K"

PN JP2002017362-A.

PD 22-JAN-2002.

PF 05-JUL-2000; 2000JP-0204236.

PR 05-JUL-2000; 2000JP-0204236.

PA (AJIN) AJINOMOTO KK.

DR MPI; 2002-263243/31.

DR P-PSDB; ABB06342.

PT Transformation of high temperature-resistant coryneform microbe and

PT transformant -

PS Example; Page 20-22; 28pp; Japanese.

CC The present invention describes a method for transforming a high

CC temperature-resistant coryneform microbe in which the high temperature-

CC resistant coryneform microbe is treated with a chemical changing the

CC structure of cell wall of the microbe and an electric pulse is applied

CC to a solution containing the microbe having a changed structure of cell

CC wall and a DNA. Also described are: (1) transforming a high temperature-

CC resistant coryneform microbe in which the cell wall of a high

CC temperature-resistant coryneform microbe is weakened and electric pulse

CC is applied to a solution containing the microbe having weakened cell

CC wall and a DNA; (2) a transformant of a high temperature-resistant

CC coryneform microbe in which a recombinant DNA is introduced; and

CC (3) a transformant of a high temperature-resistant coryneform microbe

CC carrying both a plasmid vector derived from the high temperature-

CC resistant coryneform microbe and a plasmid vector derived from

CC Corynebacterium glutamicum or Brevibacterium lactofermentum. The method

CC is used for transforming a high temperature-resistant coryneform microbe.

CC The present sequence encodes a Brevibacterium lactofermentum p48K

CC protein sequence, which is used in an example from the present

CC invention.

QY 493 AACGGCTGACGCGCCACTAGCTATTCGGG 552
DB 1795 AACGGCTGACGCGCCACTAGCTATTCGGG 1054

QY 493 AACGGTGGACAGCTCCACGCTTACGCGCTACTGATGTCAGTGTGACGTGAGTGAAGAC 552
DB 1795 AACGGTTGGACACTTGACCGGACATGCTGTGTTCTTGAGATCGTCCACTGCTGACGAT 1854
QY 553 ATCCCTGCAATCCTTCTCGAGATGCGATGCTGCAATCGATTCGATCTG 612
DB 1855 GAACCTCAAGGCGCTTTGAAGATTCATGTTTCCCGCTGCTGCTGTGTTAAGGCC 1914
QY 613 GGATTTGCTGGCGCCACTACGTAATTCGGGTG 643
DB 1915 GGTATGACGCGCCACTGCTGTGAGCACGGGG 1945

RESULT 10

AAD22582
ID AAD22582 standard; DNA; 4447 BP.

XX AAD22582;

DT 12-FEB-2002 (first entry)

DE B. lactofermentum pSFK6 replication control region DNA.

KM Coryneform bacteria; arginine repressor; L-arginine; liver function;
KW amino acid infusion; ds.

OS Brevibacterium lactofermentum.

Key Location/Qualifiers

FT CDS 1318..2598

FT /tag= a
FT /product= "B. lactofermentum pSFK6 replication control
FT region Protein"

PN BP1154020-A2.

PD 14-NOV-2001.

PF 24-APR-2001; 2001BP-0109457.

PR 28-APR-2000; 2000UP-0129167.

PA (AJIN) AJINOMOTO KK.

PI Suga M, Asakura Y, Mori Y, Ito H, Kurahashi O;

DR WPI: 2002-012660/02.

DR P-PSDB; AAB13530.

PT Coryneform bacterium with inactivated arginine repressor, useful for
PT fermentative production of arginine at high levels -

PS Example 1; Page 12-15; 30pp; English.

XX The present invention relates to Coryneform bacterium in which the
XX arginine repressor does not function normally and which produces
XX L-arginine. Coryneform bacterium is used to produce L-arginine which is
XX useful in agents for promoting liver function and in amino acid infusions
XX or in 'comprehensive' amino acid pharmaceutical. Coryneform bacterium
XX accumulate L-arginine to higher levels than wild-type strains, e.g. 120
XX compared with 20, mg/dl. The present sequence is
XX Brevibacterium lactofermentum pSFK6 replication control region DNA.
XX This DNA is used in the construction of shuttle vector for
XX Escherichia coli and Coryneform bacteria and temperature sensitive
XX vector.

SQ Sequence 4447 BP; 824 A; 997 C; 1378 G; 1248 T; 0 other;

Query Match 3.3%; Score 37.4; DB 24; Length 4447;

Best Local Similarity 53.0%; Pred. No. 0.87; Mismatches 0; Gaps 0;

Matches 80; Conservative 0; Indels 71; Indels 0; Gaps 0;

QY 493 AACGGTGGACAGCTCCACGCTTACGCGCTACTGATGTCAGTGTGACGTGAGTGAAGAC 552

DB 1795 AACGGTTGGACACTTGACCGGACATGCTGTGTTCTTGAGATCGTCCACTGCTGACGAT 1854
QY 553 ATCCCTGCAATCCTTCTCGAGATGCGATGCTGCAATCGATTCGATCTG 612
DB 1855 GAACCTCAAGGCGCTTTGAAGATTCATGTTTCCCGCTGCTGCTGTGTTAAGGCC 1914
QY 613 GGATTTGCTGGCGCCACTACGTAATTCGGGTG 643
DB 1915 GGTATGACGCGCCACTGCTGTGAGCACGGGG 1945

RESULT 11

AAD22583
ID AAD22583 standard; DNA; 4447 BP.

XX AAD22583;

DT 12-FEB-2002 (first entry)

DE B. lactofermentum p48K replication control region DNA.

KM Coryneform bacteria; arginine repressor; L-arginine; liver function;
KW amino acid infusion; mutant; ds.

OS Brevibacterium lactofermentum.

Key Location/Qualifiers

FT CDS 1318..2598

FT /tag= a
FT /product= "B. lactofermentum pSFK6 replication control
FT region Protein"

FT mutation replace (1255, C)

FT mutation replace (1534, C)

FT mutation /tag= C
FT /note= "This mutation causes an amino acid change
FT from proline to serine"

FT mutation replace (1866, G)

FT mutation /tag= d

FT mutation replace (2058, G)

FT mutation /tag= e

FT mutation replace (2187, C)

FT mutation /tag= f

FT mutation replace (3193, G)

PN BP1154020-A2.

PD 14-NOV-2001.

PF 24-APR-2001; 2001BP-0109457.

PR 28-APR-2000; 2000UP-0129167.

PA (AJIN) AJINOMOTO KK.

PI Suga M, Asakura Y, Mori Y, Ito H, Kurahashi O;

DR WPI: 2002-012660/02.

DR P-PSDB; AAB13531.

PT Coryneform bacterium with inactivated arginine repressor, useful for
PT fermentative production of arginine at high levels -

PS Example 1; Page 17-20; 30pp; English.

XX The present invention relates to Coryneform bacterium in which the
XX arginine repressor does not function normally and which produces
XX L-arginine. Coryneform bacterium is used to produce L-arginine which is
XX useful in agents for promoting liver function and in amino acid infusions
XX or in 'comprehensive' amino acid pharmaceutical. Coryneform bacterium
XX accumulate L-arginine to higher levels than wild-type strains, e.g. 120

CC compared with 20, mg/dl. The present sequence is
 CC Brevibacterium lactofermentum p48K replication control region DNA.
 CC This DNA is used in the construction of shuttle vector for
 CC Escherichia coli and Corynebacterium bacteria and temperature sensitive
 CC vector.
 XX
 SQ Sequence 4447 BP; 827 A; 993 C; 1375 G; 1252 T; 0 other;
 Query Match 3.3%; Score 37.4; DB 24; Length 4447;
 Best Local Similarity 53.0%; Pred. No. 0.87;
 Matches 80; Conservative 0; Mismatches 71; Indels 0; Gaps 0;
 QY 493 AAGCGTGGACGCGCCAGCTTCAAGCGCTACTCAGTTCAGTGGTGAACGTGAGAC 552
 DB 1795 AACGTTGGACCTTGCACCGCAACATGCTGTTCTTGATCGTCCACTGTCTGACGAT 1854
 QY 553 ATCTCGAATCTCTCTCGATGCGATGCGATGCGTGAAGCTTCAAACTCGATCTCTG 612
 DB 1855 GAACTCAAGGATTTGAGGATTCATGTTTCCGCGTGTCTGTCTGTGTGAAGGCC 1914
 QY 613 GGATTGCTGGCGCCACTACGTAATTCGAGTG 643
 DB 1915 GGTATGACGCGCCACTGCGTGAACGCGGG 1945
 RESULT 12
 AAA66016
 ID AAA66016 standard; DNA; 1631 BP.
 AC
 XX AAA66016;
 XX
 DT 05-OCT-2000 (first entry)
 XX
 DE E. coli proliferation associated coding sequence SEQ ID NO:208.
 XX
 KM Escherichia coli; E. coli; proliferation; inhibition; screening;
 KM antimicrobial; bacterial growth; antisense therapy; antibacterial; ds.
 XX
 OS Escherichia coli.
 OS
 XX WO200044906-A2.
 XX
 PN 03-AUG-2000.
 PD
 XX 27-JAN-2000; 2000WO-US02200.
 PF
 XX 27-JAN-1999; 99US-0117405.
 PR
 XX (ELIT-) ELITRA PHARM INC.
 PA
 XX Zyskind J, Ohlsen KL, Trawick J, Forsyth RA, Froelich JM, Carr GT;
 PI Yamamoto RT, Xu HH;
 PI Yamamoto RT, Xu HH;
 XX WPI: 2000-514822/46.
 DR P-PSDB; AAB16010.
 DR
 XX Novel polynucleotides and polypeptides associated with microorganism
 PT proliferation, used to identify inhibitors of bacterial growth and
 PT proliferation, for use in antisense therapy -
 PT
 XX Claim 8; Page 162-163; 316pp; English.
 PS
 XX AAA65809 and AAA65889 and AAA66058 represent nucleotide
 CC sequences derived from Escherichia coli which inhibit E. coli
 CC proliferation. AAA65890 to AAA66055 and AAB15886 to AAB16040 represent
 CC nucleotide and protein sequences associated with E. coli proliferation.
 CC AAA66056 and AAA66057 represent primers used for sequencing E. coli
 CC proliferation inhibiting nucleotide inserts in an example from the
 CC present invention. Methods for the present invention can be used to
 CC identify a proliferation- required gene in a microorganism, by contacting
 CC a microorganism with a proliferation- required gene activity inhibitory
 CC nucleic acid identified in another organism, and determining if
 CC inhibition occurs in the second microorganism. The nucleic acid sequences

CC identified as being required for bacterial growth and proliferation, can
 CC be used for antisense therapy for killing bacteria.
 XX
 SQ Sequence 1631 BP; 463 A; 340 C; 429 G; 399 T; 0 other;
 Query Match 3.1%; Score 34.8; DB 21; Length 1631;
 Best Local Similarity 50.0%; Pred. No. 3.5;
 Matches 87; Conservative 0; Mismatches 87; Indels 0; Gaps 0;
 QY 219 CGAAGAGGCTGATCTGCCCTCTGTGCGGAGAAAGTCGTGCACATGTCAGACGA 278
 DB 1184 CGAAGAGTCTGAGAGGACCGAGATGGCAAAAAAATGCGCGAAGCTATGCTGCC 1243
 QY 279 AATTTCTCAAGTTTGCTCATCACTCGGACTGATCTTTGCGATGTAAGATGAC 338
 DB 1244 ATTATCAACAAGAGTAATCTCAATTGCTGCTCAATTAATGAACAGGCGGAACTAT 1303
 QY 339 CATGGCGCATACAGCTGTGTGAGGGCTCCACGACCTATGATGATGATTTCCGC 392
 DB 1304 AGCCCGTATGAGCTGTTAAGATGCGCAAGAAATCTGTGAATTTCTGC 1357
 RESULT 13
 AAH24065/c
 ID AAH24065 standard; DNA; 4590 BP.
 AC
 XX AAH24065;
 XX
 DT 29-AUG-2001 (first entry)
 XX
 DE Yeast AOD9604-associated DNA sequence, SEQ ID NO:1.
 XX
 KM Human growth hormone analogue peptide; hGH; AOD9604; lipid metabolism;
 KM modulation; lipolysis stimulation; hormone-sensitive lipase stimulation;
 KM lipogenesis inhibition; acetyl CoA carboxylase inhibition; obesity;
 KM functional food; transgenic yeast; fat/lean ratio; food use; ds.
 XX
 OS Saccharomyces cerevisiae.
 OS
 XX
 FH Key
 FH Location/Qualifiers
 FT 10
 FT misc_feature
 FT /tag= a
 FT /note= "Represented as * in the specification"
 FT 3617
 FT misc_feature
 FT /tag= b
 FT /note= "Represented as * in the specification"
 FT 3649
 FT misc_feature
 FT /tag= c
 FT /note= "Represented as * in the specification"
 FT 3679
 FT misc_feature
 FT /tag= d
 FT /note= "Represented as * in the specification"
 FT 3819
 FT misc_feature
 FT /tag= e
 FT /note= "Represented as * in the specification"
 FT 3862
 FT misc_feature
 FT /tag= f
 FT /note= "Represented as * in the specification"
 FT 3864
 FT misc_feature
 FT /tag= g
 FT /note= "Represented as * in the specification"
 FT 3888
 FT misc_feature
 FT /tag= h
 FT /note= "Represented as * in the specification"
 FT 3890
 FT misc_feature
 FT /tag= i
 FT /note= "Represented as * in the specification"
 FT 3912
 FT misc_feature
 FT /tag= j
 FT /note= "Represented as * in the specification"
 FT 3914
 FT misc_feature
 FT /tag= k
 FT /note= "Represented as * in the specification"

```
FT misc_feature 3938
FT /*tag= 1
FT /note= "Represented as * in the specification"
FT misc_feature 3939
FT /*tag= m
FT /note= "Represented as * in the specification"
FT misc_feature 3941
FT /*tag= o
FT /note= "Represented as * in the specification"
FT misc_feature 3943
FT /*tag= p
FT /note= "Represented as * in the specification"
FT misc_feature 4361
FT /*tag= q
FT /note= "Represented as * in the specification"
```

XX WO200133977-A1.

XX 17-MAY-2001.

XX 06-NOV-2000; 2000WO-AU01362.

XX 05-NOV-1999; 99AU-0003875.

XX (META-) METABOLIC PHARM LTD.

XX Belyea CI, Ng FM, Vaughan P;

XX WPI; 2001-328876/34.

XX New organisms containing nucleic acid encoding a growth hormone
PT fragment which modulates lipid metabolism are useful to produce dietary
PT aids for obesity and in the meat production industry

XX Disclosure; Page 48-50; 54pp; English.

XX The invention relates to novel transgenic organisms useful in the
CC production of functional food and drink products for the treatment
CC or prevention of obesity via the regulation of lipid metabolism. The
CC organisms comprise a polynucleotide encoding a growth hormone fragment
CC capable of stimulating the activity of hormone-sensitive lipase (the key
CC enzyme in lipolysis) and inhibiting acetyl CoA carboxylase (the key
CC enzyme in lipogenesis). The growth hormone fragment preferably contains
CC at least the disulphide-bonded loop of a mammalian growth hormone (but is
CC not the full-length growth hormone) and is optionally linked to an
CC epitope tag or heterologous fusion protein partner. The transgenic
CC organism may be a microorganism used to produce a fermented product
CC (e.g., yeast), or an edible plant or animal or cell thereof. Food or
CC drink made using methods of the invention are used to modify fat/lean
CC ratio, lipid metabolism or food use in a mammal. In particular, the food
CC or drink products may be used to treat or prevent obesity, particularly
CC in humans, and may also be used to improve the fat/lean ration of
CC livestock raised for meat production. In the exemplification of the
CC invention, the human growth hormone (hGH) fragment analogue AOD9604 was
CC expressed in yeast, optionally fused to the FLAG epitope (AAB73625).
CC The present sequence is described as a DNA sequence from yeast in
CC the sequence listing, but is not further referred to in the
CC specification.

XX Sequence 4590 BP; 661 A; 384 C; 127 G; 522 T; 2896 other;

XX Query Match 3.1%; Score 34.8; DB 22; Length 4590;
XX Best Local Similarity 13.2%; Pred. No. 5.9;
XX Matches 75; Conservative 211; Mismatches 201; Indels 2; Gaps 1;

QY 455 ACGGATACGTGCGGCTGTGAATCACTACGAGAAAACGCTGGACGCTCCAGCTTC 514

Db 765 AMGNMTDAMGKSHSTDSKASRSBYVTYDYNHTTYANGDASSTSHBATYCNCHYNTA 706

QY 515 ACGGCTACTCATGTTCACTGACGTGACGTGAGCAACATCCTCGAATCCTTCGGATG 574

Db 705 HYDYSKNSNAGNBANCGHNTYAKRYKAVYTTDGDWSATCNGNCMTGAHSTGCTYKHK 646

QY 575 CGATGTTGATCGGNGAGCTTCCAACTGTAATCTCGGAAATTGCTGGCCACTACGTA 634
Db 645 YSNYNAGAYAVYTTDDMSATCNGMYASSHKYNHTTKDQSHKTKNTNTSNAGNBANCGH 586
QY 635 ATTCCGGTGTCTCCGATGTAACGAAAGATGGCGGTAACTGATCAAGTTCTGCTCGCT 694
Db 585 NTYAKNGHATYKTY--HBTAVYTTDDMSATCNGKTNHYGSHSTRTYVANCNNTGTAN 528
QY 695 ATCTGACGAAATTTGATGCTGGCTGTGATGAGAGTTGGTATGGCGGAGGAAAAGTG 754
Db 527 KHTDAMHATYASKNTNMSTAHGTRSBVTNAGTYKAVYTTDDMSATCNGSTWSTGT 468
QY 755 GTCCACATGCAACCGTGACCCCTGGAAATCGCTGTGATGACGATGCGGCGGATCCAC 814
Db 467 SSTNAGSTNSTRYASRSRTBTNTYNSRYASRSRTTBYBASRSRTNSTYBASRSRYAS 408
QY 815 AAGCGTTGAACTGTGCGGCGGAAATTTGATGCTGTGATGAGAGCTGGCGCAATCGCT 874
Db 407 RSNSTYRASRYASRSNSTSRNSTYRAVRANSTSTHVNANTBGTNSTYRAVSNTY 348
QY 875 GGTCCCGTGAATGCGGCGCGAGCTGTCTTGGGCGAACAATAAGATGCTCAGATCG 934
Db 347 NSTYNSSTRYASNSTYNSSTRYASNSTYNSSTTHNYRAYNSTSHGNTKTA 288
QY 935 TTGAGCAGAAAGAAATCTGCCCGTCATGCTTGCATTCATTCGCGCGCATCGATGTA 994
Db 287 TDTGVYHRTDHTSYTAYASNTTTRNDKTHSYDAYADYRAHYRASNYSHSHSNY 228
QY 995 TGATTCGACTTGTGCGCTTACGTCTTC 1023
Db 227 SRTTBSRTBTTSRTTSRTTSRTTB 199

RESULT 14

AAQ55750 ID AAQ55750 standard; DNA; 1505 BP.

AC AAQ55750;

DT 25-MAR-2003 (updated)

DT 23-JUL-1994 (first entry)

XX Genomic clone g11f includes a Brassica root specific promoter.

XX Plants; transformation; food crops; toxins; insects; fungal;

XX infection; ss.

XX Brassica napus.

OS Brassica napus.

XX Key

FT misc_feature 552..554

FT /*tag= a

FT /*note= "translation start site"

FT CAAAT_signal 438..440

FT /*tag= b

FT TATA_signal 485..488

FT /*tag= c

XX WO9402619-A1.

XX 03-FEB-1994.

XX 12-JUL-1993; 93WO-US06541.

XX 16-JUL-1992; 92US-0915246.

XX (PION-) PIONEER HI-BRED INT INC.

XX Baszczynski CJ, Bellemare G, Boivin R, Fallis L;

XX WPI; 1994-048877/06.

XX Brassica root-specific promoter sequence and vectors containing

PT it - for expression of e.g. toxins in the roots of crops
PT susceptible to root diseases
XX Disclosure; Fig 4; 54pp; English.
XX
CC The sequence shows the G1F genomic clone isolated from DNA from
CC the roots of Brassica napus and containing a promoter sequence
CC upstream from the translation start site. The promoter enhances
CC gene expression in roots and can be used to confer disease
CC resistance or immunity upon a susceptible plant root by transforming
CC the plant with a composite gene capable of abundant expression in the
CC gene, comprising the promoter and the gene of interest. This method
CC can be used to transform plants, partic food crops that have roots
CC susceptible to fungal or insect diseases. The roots of the transformed
CC plants express a toxin or substance that provides immunity or resistance
CC to the fungal or insect diseases.
CC See also AAO55749 and AAO56913-5.
CC (Updated on 25-MAR-2003 to correct PN field.)
XX
SQ Sequence 1505 BP; 395 A; 201 C; 528 G; 381 T; 0 other;
Query Match 3.0%; Score 34.6; DB 15; Length 1505;
Best Local Similarity 58.1%; Pred. No. 3.9; Indels 0; Gaps 0;
Matches 61; Conservative 0; Mismatches 44;
QY 706 ATTGCATCTGCGCTGTGATGAGGTTGAGTGGCGACGAAAAGTGGTCGACATGGC 765
DB 699 ATTGAGTTGATGATCTGTGTGTGAGGTTGGCGGAGAAAGAGTGTGTGTGT 758
QY 766 AACCGTCACCCCTGGGAAATCGTCTTTGATGACATGGCGCGGAT 810
DB 759 TACGGTGGAGCTGAAAGCATTTGTGAGAGAGAGGCGGTGGACAT 803
RESULT 15
ABI99802/c
ID ABI99802 standard; cDNA; 1954 BP.
XX
AC ABI99802;
XX
DT 07-MAR-2002 (first entry)
XX
DE Mouse ischaemic condition related cDNA sequence SEQ ID NO:908.
XX
KM Mouse; ischaemia; compressive ischaemia; occlusive ischaemia;
XX vasospastic ischaemia; ischaemic condition; ischaemic disease; ss.
OS Mus musculus.
XX
PN WO200188188-A2.
XX
PD 22-NOV-2001.
XX
PF 18-MAY-2001; 2001MO-JP04192.
XX
PR 18-MAY-2000; 2000JP-0145377.
XX
PA (UNNT-) UNIV NITHON SCHOOL JURIDICAL PERSON.
XX
PI Ishikawa K, Asai S, Takahashi Y, Nagata T, Ishii Y;
XX
DR WPI; 2002-034733/04.
XX
DR P-PDB; ABB57327.
XX
PT Examining the ischemic condition (e.g. occlusive ischemia) by measuring
PT expression levels of particular genes defined in the specification or
PT by determining the expression profile of a gene group comprising these
PT genes -
XX
PS Claim 2; Page 2294-2297; 2690pp; English.
XX
CC The present invention describes a method for examining ischaemic
CC conditions, comprising measuring the expression levels of particular

CC genes (I) in a test sample or determining the expression profile of a
CC gene group in the sample comprising genes selected from (I). The method
CC is useful for examining the ischaemic condition (e.g. compressive
CC ischaemia, occlusive ischaemia or vasospastic ischaemia) by measuring
CC expression levels of particular genes (ABI99202 to ABI99912, encoding
CC the protein sequences in ABB57020 to ABB57374) or by determining the
CC expression profile of a gene group comprising these genes. The
CC expression levels or expression profiles produced by these genes are
CC used as an indicator when screening for ischaemic condition-improving
CC drugs or therapeutics for ischaemic diseases. ABI99913 and ABI99914
CC represent PCR primers for a mouse ischaemic condition related sequence,
CC which are used in the exemplification of the present invention.
XX
SQ Sequence 1954 BP; 451 A; 599 C; 550 G; 353 T; 1 other;
Query Match 3.0%; Score 34.6; DB 24; Length 1954;
Best Local Similarity 47.8%; Pred. No. 4.5; Indels 0; Gaps 0;
Matches 100; Conservative 0; Mismatches 109;
QY 609 TCTGGATTGCTGCGCCACTACGTAATTGGGTGCTCATGTACGAAAGATCGCG 668
DB 1213 TGTGGGTGCTGTGCGCGGAGAGATGAGATGCGGATGCGGATGAGATGAGAC 1154
QY 669 TGAAGCTATCAAGTTCTGCGTGCCTATCTGACGAAAATTGATCTGGCGTTGATGA 728
DB 1153 TGTGGGCTTGGGCTTGGGCTGAGGTTGGGCTGAGGTTGGGCTTGGGCTTGA 1094
QY 729 GGTGGTATGTCGACGAGAAAAGTGTGACATGCGCAACCTGCAACCTGGGAAATCGC 788
DB 1093 GGTGTGATGTGGGCTGGGGCTGGGGCTGGGCTGGGCTGGGCTGGGCTGGGCTGG 1034
QY 789 TGTGATGACATGGGCGGGGATCCACAG 817
DB 1033 TGGCTGTGGCGAGGCTGTGTGACCAAG 1005

Search completed: November 8, 2003, 20:05:48
Job time : 284 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 8, 2003, 19:20:06 ; Search time 69 Seconds
(without alignments)
7292.417 Million cell updates/sec

Title: US-10-007-527A-1

Perfect score: 1140
Sequence: 1 atgaccagcgaagtcgctga.....taatcgcgtcgcgaatrga 1140

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents NA:*
1: /cgn2_6/prodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/prodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/prodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/prodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/prodata/2/ina/6CTUS_COMB.seq:*
6: /cgn2_6/prodata/2/ina/backfile1.seq:*

Pred. No. is the number of results predicted by chance a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	37.4	3.3	4447	US-09-521-668B-17	Sequence 17, Appl
2	37.4	3.3	4447	US-09-521-668B-19	Sequence 19, Appl
3	34.6	3.0	1505	US-07-915-246-1	Sequence 1, Appl
4	34.4	3.0	7218	US-08-232-463-14	Sequence 14, Appl
5	33.6	2.9	1134	US-09-482-180A-3	Sequence 3, Appl
6	33.2	2.9	1780	US-08-813-150-3	Sequence 3, Appl
7	33.2	2.9	1780	US-09-546-553-3	Sequence 3, Appl
8	33.2	2.9	1867	US-09-482-273-81	Sequence 81, Appl
9	32.8	2.9	1987	US-08-722-001-26	Sequence 26, Appl
10	32.8	2.9	2793	US-08-209-747-1	Sequence 1, Appl
11	32.8	2.9	2793	US-08-458-298-1	Sequence 1, Appl
12	31.6	2.8	1356	US-09-252-991A-14918	Sequence 14918, A
13	31.4	2.8	4403765	US-09-103-840A-2	Sequence 2, Appl
14	31.4	2.8	4411529	US-09-103-840A-1	Sequence 1, Appl
15	31.2	2.7	1599	US-09-252-991A-13535	Sequence 13535, A
16	31.2	2.7	2280	US-09-252-991A-13535	Sequence 13535, A
17	31.2	2.7	5741	US-07-706-699-4	Sequence 4, Appl
18	31.2	2.7	5741	US-07-998-931-4	Sequence 4, Appl
19	30.8	2.7	5820	US-09-029-213B-7	Sequence 7, Appl
20	30.6	2.7	1708	FCT-US95-02481-3	Sequence 3, Appl
21	30.6	2.7	4403765	US-09-103-840A-2	Sequence 2, Appl
22	30.6	2.7	4411529	US-09-103-840A-1	Sequence 1, Appl
23	30.4	2.7	1001	US-09-641-638-315	Sequence 15, App
24	30.4	2.7	2128	US-09-280-116-11	Sequence 11, Appl
25	30.4	2.7	2459	US-09-443-795-2	Sequence 2, Appl
26	30.4	2.7	3353	US-09-532-803-10	Sequence 10, Appl
27	30.4	2.7	3353	US-09-653-403-1	Sequence 1, Appl

ALIGNMENTS

28	30.4	2.7	3333	US-10-013-784-1	Sequence 1, Appl
29	30.2	2.6	11707	US-09-136-574A-1	Sequence 1, Appl
30	30.2	2.6	18596	US-09-318-448-11	Sequence 11, Appl
31	30.2	2.6	18597	US-09-962-665-8	Sequence 8, Appl
32	30	2.6	289	US-09-007-005-17	Sequence 17, Appl
33	30	2.6	289	US-09-244-796-17	Sequence 17, A
34	30	2.6	504	US-09-252-991A-12357	Sequence 12357, A
35	30	2.6	1200	US-09-252-991A-12403	Sequence 12403, A
36	30	2.6	1232	US-09-462-645C-3	Sequence 3, Appl
37	30	2.6	1233	US-09-252-991A-12188	Sequence 12188, A
38	30	2.6	2949	US-09-023-905A-8	Sequence 8, Appl
39	30	2.6	3481	US-09-462-645C-11	Sequence 11, Appl
40	30	2.6	3810	US-09-462-645C-7	Sequence 7, Appl
41	30	2.6	5954	US-09-023-905A-6	Sequence 6, Appl
42	30	2.6	36470	US-08-311-731A-123	Sequence 123, App
43	29.8	2.6	1353	US-09-252-991A-9189	Sequence 9189, Ap
44	29.8	2.6	1365	US-09-252-991A-13442	Sequence 13442, A
45	29.8	2.6	1464	US-09-252-991A-8876	Sequence 8876, Ap

RESULT 1

US-09-521-668B-17
; Sequence 17, Application US/09521668B
; Patent No. 6303383
; GENERAL INFORMATION:
; APPLICANT: NAKAMURA, JUN
; APPLICANT: KANNO, SOHEI
; APPLICANT: KIMURA, EICHIRO
; APPLICANT: MATSUI, KAHITO
; APPLICANT: NAKAMATSU, TOSIOHSHI
; TITLE OF INVENTION: TEMPERATURE SENSITIVE PLASMID FOR CORYNEFORM BACTERIA
; FILE REFERENCE: 0010-1093-0
; CURRENT APPLICATION NUMBER: US/09/521,668B
; CURRENT FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: JP 11-69896
; PRIOR FILING DATE: 1999-03-16
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 17
; LENGTH: 4447
; TYPE: DNA
; ORGANISM: Brevibacterium lactofermentum
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1316)..(2598)
US-09-521-668B-17

Query Match

Best Local Similarity 53.0%; Pred. No. 0.18;
Matches 80; Conservative 0; Mismatches 71; Indels 0; Gaps 0;

QY 493 AACGGTCGACGTCACGTTACGCGCTACGTCATGTCAGTCGTCAGTCGAGC 552
DB 1795 AACGGTTGGACATTCGACGCAACGATGTTGTTGATCGTCGACTGTCGACAT 1854
QY 553 ATCCGTGATTCCTTCGATGCGATGATGTCGATGCGTTCGCAACTCGATATCTG 612
DB 1855 GAACCTAAGCCGTTTGAAGATTCATGTTTCCCGGTGCTGCTGTGTTAAGGCC 1914
QY 613 GGATTTGTCGCGCCACTACGTAATTCGGGTG 643
DB 1915 GGTATGACGCGCCACTGCGTGAACGCGGG 1945
RESULT 2
US-09-521-668B-19
; Sequence 19, Application US/09521668B
; Patent No. 6303383
; GENERAL INFORMATION:
; APPLICANT: NAKAMURA, JUN

APPLICANT: KANNO, SOHEI
APPLICANT: KIMURA, EIICHIRO
APPLICANT: MATSUI, KAUIKO
APPLICANT: NAKAMATSU, TSUYOSHI
TITLE OF INVENTION: TEMPERATURE SENSITIVE PLASMID FOR CORYNEFORM BACTERIA
FILE REFERENCE: 0010-1093-0
CURRENT APPLICATION NUMBER: US/09/521,668B
CURRENT FILING DATE: 2000-03-08
PRIOR APPLICATION NUMBER: JP 11-69896
PRIOR FILING DATE: 1999-03-16
NUMBER OF SEQ ID NOS: 20
SOFTWARE: PatentIn version 3.0
SEQ ID NO 19
LENGTH: 4447
TYPE: DNA
ORGANISM: Brevibacterium lactofermentum
FEATURE:
NAME/KEY: CDS
LOCATION: (1318)..(2598)
US-09-521-668B-19

Query Match 3.3%; Score 37.4; DB 4; Length 4447;
Best Local Similarity 53.0%; Pred. No. 0.18; Indels 0; Gaps 0;
Matches 80; Conservative 0; Mismatches 71; Indels 0; Gaps 0;

QY 493 AACGGCTGGACCGTCCACGTTTCAGCGCTACTCATGTTCACTGAGTGAAGAAC 552
DB 1795 AACGGTTGGACCTTGACCGCAACATGCTGTGTTCTTGATGTCACATGTTGAGAT 1854
QY 553 ATCCTGATCTCTTCTCGATGCGATGCTGATGCGTGAAGTTCGAACTGATCTCTG 612
DB 1855 GAACCTAAGGCACTTTGAGATTCATGTTTCCCGCTGCTCTGCTGTTAAAGGCC 1914
QY 613 GGATTGCTGGCGCCACAGCTAAATTCGGGTG 643
DB 1915 GGATGAGCGCGCCACTGCGTGAAGCAGCGG 1945

RESULT 3

US-07-915-246-1
Sequence 1, Application US/07915246
Patent No. 5401836

GENERAL INFORMATION:

APPLICANT: Baszczyński, Chris L.

APPLICANT: Fallis, Lynne

APPLICANT: Bellmare, Guy

APPLICANT: Boivin, Rodolphe

TITLE OF INVENTION: A BRASSICA REGULATORY SEQUENCE FOR

NUMBER OF SEQUENCES: 5 ROOT-SPECIFIC OR ROOT-ABUNDANT GENE EXPRESSION

CORRESPONDENCE ADDRESS:

ADDRESSEE: McAndrews, Held, and Malloy

STREET: 500 W. Madison St. 34th Floor

CITY: Chicago

STATE: Illinois

COUNTRY: USA

ZIP: 60661

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/07/915,246

FILING DATE: 19920716

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Pochopien, Donald J.

REGISTRATION NUMBER: 32,167

REFERENCE/DOCKET NUMBER: 91 P 1125

TELECOMMUNICATION INFORMATION:

TELEPHONE: 312 707-8889

TELEFAX: 312 707-9155

INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1505 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: double
TOPOLOGY: linear
MOLSCULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Brassica napus
STRAIN: Westar
DEVELOPMENTAL STAGE: Somatic
TISSUE TYPE: Root
US-07-915-246-1

Query Match 3.0%; Score 34.6; DB 1; Length 1505;
Best Local Similarity 58.1%; Pred. No. 0.8;
Matches 61; Conservative 0; Mismatches 44; Indels 0; Gaps 0;

QY 706 ATTGATCTGCGCTTGATGAGGTTGTGAGCGCAAGAAAAAGTGTGACATGCG 765
DB 699 ATTGAGTTGATGCTGTGTTGTGTTGAGTGTGTGGAGAGAGAGAGTGTGTGTGT 758
QY 766 AACCGTGACCCCTGGAAATCGCTGTTGATGATGAGTGGCGCGGAT 810
DB 759 TAGCGTGAGCTGAAGGACATGTGTGAGAGAGAGCGGTGACAT 803

RESULT 4

US-08-232-463-14/c
Sequence 14, Application US/08232463
Patent No. 5670367

GENERAL INFORMATION:

APPLICANT: DORNER, F.

APPLICANT: SCHRIFFLINGER, F.

TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS

NUMBER OF SEQUENCES: 52

CORRESPONDENCE ADDRESS:

ADDRESSEE: Foley & Lardner

STREET: 1800 Diagonal Road, Suite 500

CITY: Alexandria

STATE: VA

COUNTRY: USA

ZIP: 22313-0299

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/232,463

FILING DATE: 20000716

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/07/935,313

FILING DATE:

APPLICATION NUMBER: EP 91 114 300.6

ATTORNEY/AGENT INFORMATION:

NAME: BENT, Stephen A.

REGISTRATION NUMBER: 29,768

REFERENCE/DOCKET NUMBER: 30472/114 INMU

TELECOMMUNICATION INFORMATION:

TELEPHONE: (703)836-9300

TELEFAX: (703)836-4109

TELEX: 899149

INFORMATION FOR SEQ ID NO: 14:

SEQUENCE CHARACTERISTICS:

LENGTH: 7218 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

417 TCGTCGTTGGCCGTACGGAAAGTAAATGTACGGCTGCAGCGAATACGTGCGCGCTGTGA 476

DB 1016 GGTCGTCTGACCCGAGCCGACCGACCGCCGTCGATGAAGCGCAGTCACTTCCTCGCCGCGCTGCC 1075

QY 700 ACGAAATATGCATCTGCGCGTGTGATGAGAGTTGGTAGTGGCGACGAAAAAGTGTGC 757

Db 1076 CCGCCACATGACAGAGGCGCTGCGAGGCTGCACTACGTGCGCGCCACCAAGGTGTTCC 1133

RESULT 7

US-09-546-553-3
Sequence 3, Application US/09546553
Patent No. 6518405

GENERAL INFORMATION:

APPLICANT: Mueller, Christopher
APPLICANT: Lebecque, Serge J.E.
APPLICANT: Liu, Yong-Jun
APPLICANT: Dowling, Lynette M.
APPLICANT: Huffine, Constance M.
APPLICANT: Gorman, Daniel M.
TITLE OF INVENTION: MAMMALIAN PROTEINASES; OXIDOREDUCTASES;
TITLE OF INVENTION: RELATED REAGENTS
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSER: DNAX Research Institute
STREET: 901 California Avenue
City: Palo Alto
STATE: California
COUNTRY: USA
ZIP: 94304-1104

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/546,553
FILING DATE: 10-APR-2000
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/813,150
FILING DATE: 07-MAR-1997
ATTORNEY/AGENT INFORMATION:
NAME: Chang, Edwin P.
REGISTRATION NUMBER: 34,090
REFERENCE/DOCKET NUMBER: SF0693
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-852-9196
TELEFAX: 650-496-1200
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 1780 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 66..1766
US-09-546-553-3

US-09-546-553-3

Query Match 2.9%; Score 33.2; DB 4; Length 1780;
Best Local Similarity 55.1%; Pred. No. 2.4;
Matches 65; Conservative 0; Mismatches 53; Indels 0; Gaps 0;

QY 640 GGTGCTCGATGTACGAAAGATCGCGGTGAAGCTGATCAAGTTCTGCGGTATCTG 699
DB 1016 GGTGCTGCTGACGCGGACCGGACCGGCGGTGAAGCGCATCACTTCTGCGCGCGTGC 1075
QY 700 ACGAAATTCATCTGCGGTGTGTATGAGGTTGTAGTGGCGACGGAAGAGTGC 757
DB 1076 CCGCCACATGACAGAGGCGCTGCGAGGCTGCACTACGTGCGCGCCACCAAGGTGTTCC 1133

RESULT 8
US-09-482-273-81
Sequence 81, Application US/09482273
Patent No. 6534631

GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: 71 Human Secreted Proteins
FILE REFERENCE: P2030P1
CURRENT APPLICATION NUMBER: US/09/482,273
CURRENT FILING DATE: 2000-01-13
EARLIER APPLICATION NUMBER: PCT/US99/15849
EARLIER FILING DATE: 1999-07-14
EARLIER APPLICATION NUMBER: 60/092,921
EARLIER FILING DATE: 1998-07-15
EARLIER APPLICATION NUMBER: 60/092,922
EARLIER FILING DATE: 1998-07-15
EARLIER APPLICATION NUMBER: 60/092,956
EARLIER FILING DATE: 1998-07-15
NUMBER OF SEQ ID NOS: 267
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 81
LENGTH: 1867
TYPE: DNA
ORGANISM: Homo sapiens
US-09-482-273-81

Query Match 2.9%; Score 33.2; DB 4; Length 1867;
Best Local Similarity 55.1%; Pred. No. 2.5;
Matches 65; Conservative 0; Mismatches 53; Indels 0; Gaps 0;

QY 640 GGTGCTCGATGTACGAAAGATCGCGGTGAAGCTGATCAAGTTCTGCGGTATCTG 699
DB 1053 GGTGCTGCTGACGCGGACCGGACCGGCGGTGAAGCGCATCACTTCTGCGCGCGTGC 1112
QY 700 ACGAAATTCATCTGCGGTGTGTATGAGGTTGTAGTGGCGACGGAAGAGTGC 757
DB 1113 CCGCCACATGACAGAGGCGCTGCGAGGCTGCACTACGTGCGCGCCACCAAGGTGTTCC 1170

RESULT 9

US-08-722-001-26
Sequence 26, Application US/08722001
Patent No. 5760054

GENERAL INFORMATION:

APPLICANT: Thompson, Wayne J.
APPLICANT: Huff, Joel R.
APPLICANT: Nerendery, Jennie B.
APPLICANT: Lee, Hee-Yoon
TITLE OF INVENTION: ALPHALIC ADRENERGIC RECEPTOR ANTAGONISTS
NUMBER OF SEQUENCES: 35
CORRESPONDENCE ADDRESS:
ADDRESSER: Merck & Co., Inc.
STREET: 126 Lincoln Avenue
CITY: Rahway
STATE: New Jersey
COUNTRY: United States of America
ZIP: 07065

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/722,001
FILING DATE:
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/229,276
FILING DATE: 14-APR-1995
ATTORNEY/AGENT INFORMATION:
NAME: Appollina, Mary A.
REGISTRATION NUMBER: 34,087
REFERENCE/DOCKET NUMBER: 19169Y
TELECOMMUNICATION INFORMATION:
TELEPHONE: (908) 594-3462
TELEFAX: (908) 594-4720

TELEX: 138825
INFORMATION FOR SEQ ID NO: 26;
SEQUENCE CHARACTERISTICS:
LENGTH: 1987 base pairs
TYPE: nucleic acid
STRANDEDNESS: both
TOPOLOGY: both
MOLECULE TYPE: cDNA
HYPOHETICAL: NO
ANTI-SENSE: NO
US-08-722-001-26

Query Match 2.9%; Score 32.8; DB 1; Length 1987;
Best Local Similarity 56.5%; Pred. No. 3.4;
Matches 61; Conservative 0; Mismatches 47; Indels 0; Gaps 0;

QY 105 ACAATCCACCCGCTCAAGAAACATTAAAGCGCTGTGGCCGCCGATTTTGGCTGAACG 164
DB 778 ACCGACTACTACATCGTCGACACCGCTAGTGGCGGTGGCCGACCTCCGCTCACTCCACG 837
QY 165 TGTGACCATGTTCAGAGGTCGCAAGGTTTGGATTGCGAGGCTTCG 212
DB 838 TGTGCTCCCTCTCCGCCATCTTCAGAGTCTTAGGCTACTGGGCTTCG 885

RESULT 10
US-08-209-747-1
Sequence 1, Application US/08209747
Patent No. 5733771
GENERAL INFORMATION:
APPLICANT: Lewis, Randolph V.
TITLE OF INVENTION: cDNAs Encoding Minor Ampullate Spider
TITLE OF INVENTION: Silk Proteins
NUMBER OF SEQUENCES: 56
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Birch, Stewart, Kolaesch & Birch
STREET: P.O. Box 747
CITY: Falls Church
STATE: Virginia
COUNTRY: USA
ZIP: 22040-3487
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/209,747
FILING DATE: 14-MAR-1994
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Murphy Jr., Gerald M.
REGISTRATION NUMBER: 28,977
REFERENCE/DOCKET NUMBER: 1447-104P
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-205-8000
TELEFAX: 703-205-8050
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2793 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOHETICAL: NO
ORIGINAL SOURCE:
ORGANISM: Nephila clavipes
TISSUE TYPE: minor ampullate gland
FEATURE:
NAME/KEY: CDS
LOCATION: 183..2675
OTHER INFORMATION: /product= "N. clavipes minor

OTHER INFORMATION: ampullate silk protein"
US-08-209-747-1

Query Match 2.9%; Score 32.8; DB 1; Length 2793;
Best Local Similarity 39.8%; Pred. No. 4.1;
Matches 82; Conservative 0; Mismatches 124; Indels 0; Gaps 0;

QY 708 TGCATGCGCGTTGGTATGAGGTTGTGTGCGGACGAAAAAGTGTGACATGCGAA 767
DB 899 TGGAGCTGAGCGCGCTGACGAGTGACAGACGAGAGCTGAGAGCTACGCTGTCAAGATA 958
QY 768 CCGTGACCCCTGGGAAATGCGCTGTGATGACAGTGGCGGAGATCCACAAGCGTTGAAC 827
DB 959 CGGTGCCGAGACGAGAGCTGTGCTGCTGCTGACGCTGCAANNNNNNNNNNNNNNNN 1018
QY 828 GTGGCAGAAATTGACTTTGTTTCATGAGACGTCGCGCATCGCTGTCCCGTGATT 887
DB 1019 NNN 1078
QY 888 GCGTGCCGAGCTGTCTTTGGGCGAG 913
DB 1079 TGGTGCCGAGCAGAGAGCTGTGCGG 1104

RESULT 11
US-08-458-298-1
Sequence 1, Application US/08458298
Patent No. 5756677
GENERAL INFORMATION:
APPLICANT: Lewis, Randolph V.
TITLE OF INVENTION: cDNAs Encoding Minor Ampullate Spider
TITLE OF INVENTION: Silk Proteins
NUMBER OF SEQUENCES: 56
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Birch, Stewart, Kolaesch & Birch
STREET: P.O. Box 747
CITY: Falls Church
STATE: Virginia
COUNTRY: USA
ZIP: 22040-3487
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/458,298
FILING DATE: 02-JUN-1995
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/209,747
FILING DATE: 14-MAR-1994
ATTORNEY/AGENT INFORMATION:
NAME: Murphy Jr., Gerald M.
REGISTRATION NUMBER: 28,977
REFERENCE/DOCKET NUMBER: 1447-104P
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-205-8000
TELEFAX: 703-205-8050
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2793 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOHETICAL: NO
ORIGINAL SOURCE:
ORGANISM: Nephila clavipes
TISSUE TYPE: minor ampullate gland
FEATURE:
NAME/KEY: CDS

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 8, 2003, 20:01:41 / Search time 304 Seconds
(without alignments)

11966.092 Million cell updates/sec

Title: US-10-007-527a-1

Perfect score: 1140
Sequence: 1 atgaccagcgttaagtcgtga.....taataatcggttcgcaagtga 1140

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 2141354 seqs, 1595478879 residues

Total number of hits satisfying chosen parameters: 4282708

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-Processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications NA.*

```
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq:*
10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq:*
11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:*
12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
13: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq:*
14: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:*
15: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
16: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
17: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1140	100.0	1140	US-10-007-527a-1	Sequence 1, Appli
2	1140	100.0	1140	US-10-007-452-1	Sequence 1, Appli
3	1140	100.0	6334	US-10-007-527a-5	Sequence 5, Appli
4	1140	100.0	6334	US-10-007-452-5	Sequence 5, Appli
5	1140	100.0	9652	US-10-007-527a-7	Sequence 7, Appli
6	1140	100.0	9652	US-10-007-452-7	Sequence 7, Appli
7	1140	100.0	11241	US-10-007-527a-6	Sequence 6, Appli
8	1140	100.0	11241	US-10-007-452-6	Sequence 6, Appli
9	40.8	3.6	520	US-10-184-644-332	Sequence 332, App
10	40.8	3.6	520	US-10-184-634-332	Sequence 332, App
11	37.6	3.3	594	US-10-140-472-10	Sequence 10, Appl
12	37.6	3.3	594	US-10-141-761-10	Sequence 10, Appl
13	37.6	3.3	594	US-10-142-885-10	Sequence 10, Appl
14	37.6	3.3	594	US-10-158-790-10	Sequence 10, Appl
15	37.6	3.3	594	US-10-123-155-10	Sequence 10, Appl
16	37.6	3.3	594	US-10-146-731-10	Sequence 10, Appl

17	37.4	3.3	4447	9	US-09-835-381-5	Sequence 5, Appli
18	37.4	3.3	4447	9	US-09-835-381-7	Sequence 7, Appli
19	37.4	3.3	4447	14	US-10-196-232-5	Sequence 5, Appli
20	37.4	3.3	4447	14	US-10-196-232-6	Sequence 6, Appli
21	36	3.2	299	14	US-10-184-644-164	Sequence 164, App
22	36	3.2	299	14	US-10-184-644-164	Sequence 164, App
23	34.8	3.1	1631	9	US-09-912-020-208	Sequence 208, App
24	34.4	3.0	999	14	US-10-184-644-434	Sequence 434, App
25	34.4	3.0	999	14	US-10-184-644-434	Sequence 434, App
26	34.2	3.0	623	12	US-10-027-632-11021	Sequence 11021, A
27	34.2	3.0	623	12	US-10-027-632-11021	Sequence 11021, A
28	34	3.0	1049	12	US-10-140-472-358	Sequence 358, App
29	34	3.0	1049	12	US-10-141-761-358	Sequence 358, App
30	34	3.0	1049	12	US-10-142-885-358	Sequence 358, App
31	34	3.0	1049	12	US-10-158-790-358	Sequence 358, App
32	34	3.0	1049	14	US-10-123-155-358	Sequence 358, App
33	34	3.0	1049	15	US-10-146-731-358	Sequence 358, App
34	33.6	2.9	1134	13	US-10-021-758-3	Sequence 3, Appli
35	33.6	2.9	1338	10	US-09-738-626-1200	Sequence 1200, App
36	33.6	2.9	1972	12	US-10-017-161-1981	Sequence 1981, App
37	33.6	2.9	3309400	10	US-09-738-626-1	Sequence 1, Appli
38	33.6	2.9	9025608	14	US-10-156-761-1	Sequence 1761, App
39	33.4	2.9	1390	12	US-10-017-161-1781	Sequence 1781, App
40	33.4	2.9	58985	11	US-09-901-152-3	Sequence 3, Appli
41	33.4	2.9	143601	12	US-09-855-824-3	Sequence 3, Appli
42	33.2	2.9	549	11	US-09-991-936-1833	Sequence 1833, App
43	33.2	2.9	1780	12	US-10-349-806-3	Sequence 3, Appli
44	33.2	2.9	1798	11	US-09-938-795A-4	Sequence 4, Appli
45	33.2	2.9	1844	11	US-09-946-374-83	Sequence 83, Appli

ALIGNMENTS

```
RESULT 1
US-10-007-527a-1
; Sequence 1, Application US/10007527a
; Publication No. US20030044807A1
; GENERAL INFORMATION:
; APPLICANT: Tomb, Jean-Francois
; APPLICANT: Bramecci, Michael G.
; APPLICANT: Cheng, Qiong
; TITLE OF INVENTION: Rhodococcus Cloning and Expression Vectors
; FILE REFERENCE: C14709 US NA
; CURRENT APPLICATION NUMBER: US/10/007,527A
; PRIOR FILING DATE: 2001-12-05
; PRIOR APPLICATION NUMBER: 60/254,868
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 1
; LENGTH: 1140
; TYPE: DNA
; ORGANISM: Rhodococcus AM12
US-10-007-527a-1
```

```
Query Match 100.0%; Score 1140; DB 14; Length 1140;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1140; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGACGAGCGTAAGTCGTGAACACCTTTCCGCAAGACCGGCTCCGTCGTGCG 60
Db 1 ATGACGAGCGTAAGTCGTGAACACCTTTCCGCAAGACCGGCTCCGTCGTGCG 60
QY 61 TCCGATTAAGCGCGGCGATCCGCGACGAACTGCGAACCACTTCAACCAACGCTGA 120
Db 61 TCCGATTAAGCGCGGCGATCCGCGACGAACTGCGAACCACTTCAACCAACGCTGA 120
QY 121 GAAACATTTAAAGCCCTGTGCGCGCGCGGATTTCTGGGGTGAAGGTGTGCAATTGCAAC 180
Db 121 GAAACATTTAAAGCCCTGTGCGCGCGCGGATTTCTGGGGTGAAGGTGTGCAATTGCAAC 180
```

```
QY 181 GGTCCGAAAGTTCTGAGATTCGAGGCTTCTGTTCTGCGAAAGGGCTGATTCGCC 240
DB 181 GGTCCGAAAGTTCTGAGATTCGAGGCTTCTGTTCTGCGAAAGGGCTGATTCGCC 240
QY 241 TGTGTGCGGGAAGAAATCGGTGACATCGTGCAGAGAAATTTCTCAAGTTGTGCTAT 300
DB 241 TGTGTGCGGGAAGAAATCGGTGACATCGTGCAGAGAAATTTCTCAAGTTGTGCTAT 300
QY 301 CAACCTGGGACTGATCTGTGCGATGTGACGATGACATGACATGACGATGACGATGAC 360
DB 301 CAACCTGGGACTGATCTGTGCGATGTGACGATGACATGACATGACGATGACGATGAC 360
QY 361 CGGCTCCACGACCTTATGACCTGACCTTTGCGAGCTTGGAAAGCTGCAACCGTCTG 420
DB 361 CGGCTCCACGACCTTATGACCTTGGAGCTTTGCGAGCTTGGAAAGCTGCAACCGTCTG 420
QY 421 CGTGGCGGTACGGAAGGTGAATGTACGCTGCGACGATGACGATGACGATGACGAT 480
DB 421 CGTGGCGGTACGGAAGGTGAATGTACGCTGCGACGATGACGATGACGATGACGAT 480
QY 481 ACTCAGGAAAAAGGCTGACGCTGACGCTGACGCTGACGCTGACGCTGACGCTGAC 540
DB 481 ACTCAGGAAAAAGGCTGACGCTGACGCTGACGCTGACGCTGACGCTGACGCTGAC 540
QY 541 GTGAGTGAAGAACCTCTGCAATCTTCTGCGATGCGATGCGATGCGATGCGATGCG 600
DB 541 GTGAGTGAAGAACCTCTGCAATCTTCTGCGATGCGATGCGATGCGATGCGATGCG 600
QY 601 CTGCTATCTCTGCGATTTGCTGCGACGATGCGATGCGATGCGATGCGATGCGATG 660
DB 601 CTGCTATCTCTGCGATTTGCTGCGACGATGCGATGCGATGCGATGCGATGCGATG 660
QY 661 ATCGGCGGTGAAGCTGATCAAGTTCTGCGCTGATGCGATGCGATGCGATGCGATG 720
DB 661 ATCGGCGGTGAAGCTGATCAAGTTCTGCGCTGATGCGATGCGATGCGATGCGATG 720
QY 721 GGTATGAGGTTGTGATGTCGACGAGAAAAAGTGTGACATGCGACGACGACGACG 780
DB 721 GGTATGAGGTTGTGATGTCGACGAGAAAAAGTGTGACATGCGACGACGACGACG 780
QY 781 GAAATGCTGTTTATGATGAGTGGGCGGAGATGCAAGCGTTGGAACGTGCGAGAAAT 840
DB 781 GAAATGCTGTTTATGATGAGTGGGCGGAGATGCAAGCGTTGGAACGTGCGAGAAAT 840
QY 841 GAGTTGTTGATGATGAGGAGCTCGGCGAATGCGGTGATGCGGTGATGCGGTGATG 900
DB 841 GAGTTGTTGATGATGAGGAGCTCGGCGAATGCGGTGATGCGGTGATGCGGTGATG 900
QY 901 GGTCTTGGGCGAGAACTTACAGATGCTCAGATGCTTGAAGAGAAATCTGCGCGT 960
DB 901 GGTCTTGGGCGAGAACTTACAGATGCTCAGATGCTTGAAGAGAAATCTGCGCGT 960
QY 961 ATGATGCGATGATCTTCCGCGGATGCTGATGATGATGATGATGATGATGATGATG 1020
DB 961 ATGATGCGATGATCTTCCGCGGATGCTGATGATGATGATGATGATGATGATGATG 1020
QY 1021 TTGGCGAGATCTCTCGGACTCGTGAAGCTGCGCGCACTTGGGAAAAATCTTCTGAT 1080
DB 1021 TTGGCGAGATCTCTCGGACTCGTGAAGCTGCGCGCACTTGGGAAAAATCTTCTGAT 1080
QY 1081 TTGCATTATGATTTGCCCGAGCGGATGTGCGGCCCCCGATATATGCGTTCCAAAG 1140
DB 1081 TTGCATTATGATTTGCCCGAGCGGATGTGCGGCCCCCGATATATGCGTTCCAAAG 1140
```

RESULT 2

```
US-10-007-452-1
; Sequence 1, Application US/10007452
; Publication No. US20030093701A1
; GENERAL INFORMATION:
; APPLICANT: Tomb, Jean-Francois
; APPLICANT: Bramucci, Michael G.
; APPLICANT: Cheng, Qiong
```

```
; APPLICANT: Kostichka, Kristy N.
; TITLE OF INVENTION: Rhodococcus Cloning and Expression Vectors
; FILE REFERENCE: CL1709 US NA
; CURRENT APPLICATION NUMBER: US/10/007,452
; CURRENT FILING DATE: 2001-11-08
; PRIOR APPLICATION NUMBER: 60/254,868
; PRIOR FILING DATE: 2000-12-12
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 1
; LENGTH: 1140
; TYPE: DNA
; ORGANISM: Rhodococcus AM12
US-10-007-452-1
```

Query Match 100.0%; Score 1140; DB 14; Length 1140;

Best Local Similarity 100.0%; Pred. No. 0; Mismatches 0; Indels 0; Gaps 0;

Matches 1140; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 ATGACGAGCGTAAGTGTGAACACCTTCCGCAAGAACCGGCTCCGCTCGTGTGCG 60
DB 1 ATGACGAGCGTAAGTGTGAACACCTTCCGCAAGAACCGGCTCCGCTCGTGTGCG 60
QY 61 TCCGATTAAGCGCGCATCCGCGACGATCGGACCCCAACTTCAACAAATCCACGTCA 120
DB 61 TCCGATTAAGCGCGCATCCGCGACGATCGGACCCCAACTTCAACAAATCCACGTCA 120
QY 121 GAAACATTTAAAGCGCTGTCGCGCGCGGATTTCTGCGGTGAACGATGTCATGTCAC 180
DB 121 GAAACATTTAAAGCGCTGTCGCGCGCGGATTTCTGCGGTGAACGATGTCATGTCAC 180
QY 181 GGTCCGAAAGTTCTGAGATTCGAGGCTTCTGTTCTGCGAAAGGGCTGATTCGCC 240
DB 181 GGTCCGAAAGTTCTGAGATTCGAGGCTTCTGTTCTGCGAAAGGGCTGATTCGCC 240
QY 241 TGTGTGCGGGAAGAAATCGGTGACATCGTGCAGAGAAATTTCTCAAGTTGTGCTAT 300
DB 241 TGTGTGCGGGAAGAAATCGGTGACATCGTGCAGAGAAATTTCTCAAGTTGTGCTAT 300
QY 301 CAACCTGGGACTGATCTGTGCGATGTGACGATGACATGACGATGACGATGACGATG 360
DB 301 CAACCTGGGACTGATCTGTGCGATGTGACGATGACATGACGATGACGATGACGATG 360
QY 361 CGGCTCCACGACCTTATGACCTTGGAGCTTGGAAAGCTGCAACCGTCTG 420
DB 361 CGGCTCCACGACCTTATGACCTTGGAGCTTGGAAAGCTGCAACCGTCTG 420
QY 421 CGTGGCGGTACGGAAGGTGAATGTACGCTGCGACGATGACGATGACGATGACG 480
DB 421 CGTGGCGGTACGGAAGGTGAATGTACGCTGCGACGATGACGATGACGATGACG 480
QY 481 ACTCAGGAAAAAGGCTGACGCTGACGCTGACGCTGACGCTGACGCTGACGCTGAC 540
DB 481 ACTCAGGAAAAAGGCTGACGCTGACGCTGACGCTGACGCTGACGCTGACGCTGAC 540
QY 541 GTGAGTGAAGAACCTCTGCAATCTTCTGCGATGCGATGCGATGCGATGCGATGCG 600
DB 541 GTGAGTGAAGAACCTCTGCAATCTTCTGCGATGCGATGCGATGCGATGCGATGCG 600
QY 601 CTGCTATCTCTGCGATTTGCTGCGACGATGCGATGCGATGCGATGCGATGCGATG 660
DB 601 CTGCTATCTCTGCGATTTGCTGCGACGATGCGATGCGATGCGATGCGATGCGATG 660
QY 661 ATCGGCGGTGAAGCTGATCAAGTTCTGCGCTGATGCGATGCGATGCGATGCGATG 720
DB 661 ATCGGCGGTGAAGCTGATCAAGTTCTGCGCTGATGCGATGCGATGCGATGCGATG 720
QY 721 GGTATGAGGTTGTGATGTCGACGAGAAAAAGTGTGACATGCGACGACGACGACG 780
DB 721 GGTATGAGGTTGTGATGTCGACGAGAAAAAGTGTGACATGCGACGACGACGACG 780
QY 781 GAAATGCTGTTTATGATGAGTGGGCGGAGATGCAAGCGTTGGAACGTGCGAGAAAT 840
DB 781 GAAATGCTGTTTATGATGAGTGGGCGGAGATGCAAGCGTTGGAACGTGCGAGAAAT 840
```

Db 781 GAATCGCTGTATGATGAGGCGGAGATCCAAAGCGTTGGAAGCTGTGGCAGAAATTT 840
QY 841 GAATTTGGTTGATGAGGAGCTGCGGCAATGCGTGTGCTCCGTGATTTGCTGCCAGCT 900
Db 841 GAGTTTGGTTGATGAGGAGCTGCGGCAATGCGTGTGCTCCGTGATTTGCTGCCAGCT 900
QY 901 GGTCTTGGGCGAAGACTTAAGATGCTCAGATGTTTGAAGAGAAATTTGCCCGGCTC 960
Db 901 GGTCTTGGGCGAAGACTTAAGATGCTCAGATGTTTGAAGAGAAATTTGCCCGGCTC 960
QY 961 ATGTTGGCGATTCATTCGCGCGCATGCTGATGATGATTTGCACTTGTGGCCTTACGTC 1020
Db 961 ATGTTGGCGATTCATTCGCGCGCATGCTGATGATGATTTGCACTTGTGGCCTTACGTC 1020
QY 1021 TTGCGGAGATTCCTCGGACTGTCGAGCTGCGGCACTTGGGAAAATCTTCGTGATCAC 1080
Db 1021 TTGCGGAGATTCCTCGGACTGTCGAGCTGCGGCACTTGGGAAAATCTTCGTGATCAC 1080
QY 1081 TTGCATTATGATTTGCCCGCAGCGATGTGCGGCCCCCGATTAATATGCTTCCCAAGTGA 1140
Db 1081 TTGCATTATGATTTGCCCGCAGCGATGTGCGGCCCCCGATTAATATGCTTCCCAAGTGA 1140

RESULT 3

US-10-007-527a-5/c
; Sequence 5, Application US/10007527A
; Publication No. US20030044807A1
; GENERAL INFORMATION:
; APPLICANT: Tomb, Jean-Francois
; APPLICANT: Bramucci, Michael G.
; APPLICANT: Cheng, Qiong
; APPLICANT: Kostichka, Kristy N.
; TITLE OF INVENTION: Rhodococcus Cloning and Expression Vectors
; FILE REFERENCE: C11709 US NA
; CURRENT APPLICATION NUMBER: US/10/007, 527A
; PRIOR FILING DATE: 2001-12-05
; PRIOR APPLICATION NUMBER: 60/254, 868
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 5
; LENGTH: 6334
; TYPE: DNA
; ORGANISM: Rhodococcus AN12
US-10-007-527a-5

Query Match 100.0%; Score 1140; DB 14; Length 6334;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1140; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGACGACGCTAAGTGTGTAACACCTTCCGCGCAAAAGACGCGCTCCCGTCCGTGTGTCG 60
Db 3051 ATGACGACGCTAAGTGTGTAACACCTTCCGCGCAAAAGACGCGCTCCCGTCCGTGTGTCG 2992
QY 61 TCCGATTAAGCGCGCATCCGCGCAGAACTGCGAACCCAACTTCAACAATACCAAGTCA 120
Db 2991 TCCGATTAAGCGCGCATCCGCGCAGAACTGCGAACCCAACTTCAACAATACCAAGTCA 2932
QY 121 GAAACATTTAACGCTGTGCGCGCCGATTTCTGCGTGAACGCTGTGACCATTTGTCAAC 180
Db 2931 GAAACATTTAACGCTGTGCGCGCCGATTTCTGCGTGAACGCTGTGACCATTTGTCAAC 2872
QY 181 GGTCCGAAAGGTTCTGATTCGAGAGGCTTGCTGCTGCGGAAAGGCTGATTCGCCCC 240
Db 2871 GGTCCGAAAGGTTCTGATTCGAGAGGCTTGCTGCTGCGGAAAGGCTGATTCGCCCC 2812
QY 241 TGCTGTGCGGAAAAAGTGTGCAATCGTGCAGACGAAATTTTCAAGTTGTGTCAT 300
Db 2811 TGCTGTGCGGAAAAAGTGTGCAATCGTGCAGACGAAATTTTCAAGTTGTGTCAT 2752
QY 301 CAACTCGGGAAGTGTGTCGATGTCGATGTCGATGTCGATGTCGATGTCGATGTCGATGTCG 360
Db 2751 CAACTCGGGAAGTGTGTCGATGTCGATGTCGATGTCGATGTCGATGTCGATGTCGATGTCG 2692

QY 361 CGGCTTCACAGACCTATGACTGACTTTTGGGAGGCTTGAAGCTGCCACCAAGCTGCT 420
Db 2691 CGGCTTCACAGACCTATGACTGACTTTTGGGAGGCTTGAAGCTGCCACCAAGCTGCT 2632
QY 421 CGTTGGCGTACGGAAGCTGAAATGTAACGCTGCGACGGAATACGTCGCGCTGTTGAATC 480
Db 2631 CGTTGGCGTACGGAAGCTGAAATGTAACGCTGCGACGGAATACGTCGCGCTGTTGAATC 2572
QY 481 ACTCAAGAAAAACGCGTGGACGTCACGTCACGCTTCACTCACTGTTCAAGTGTGAC 540
Db 2571 ACTCAAGAAAAACGCGTGGACGTCACGTCACGCTTCACTCACTGTTCAAGTGTGAC 2512
QY 541 GTGAGTGAAGAACCTCTGCAATCTTCTGATGATGATGATGATGATGATGATGATGAT 600
Db 2511 GTGAGTGAAGAACCTCTGCAATCTTCTGATGATGATGATGATGATGATGATGATGAT 2452
QY 601 CTGCTATCTGCGGATTTGCTGCGGACGACATGCAATTTGCGGATGCTGATGCAAGAA 660
Db 2451 CTGCTATCTGCGGATTTGCTGCGGACGACATGCAATTTGCGGATGCTGATGCAAGAA 2392
QY 661 ATCGGCGTGAAGCTGATCAAGTTCTGCTGCTATCTGACGAAATTTGATCTGCGCTT 720
Db 2391 ATCGGCGTGAAGCTGATCAAGTTCTGCTGCTATCTGACGAAATTTGATCTGCGCTT 2332
QY 721 GGTATGAGAGTTGTGATGTCGCGACGGAAGAAAGTGTGCAATGCGCAACCGTGTG 780
Db 2331 GGTATGAGAGTTGTGATGTCGCGACGGAAGAAAGTGTGCAATGCGCAACCGTGTG 2272
QY 781 GAATGCGCTGTGATGTCAGTGGCGGCGGATTCACAGCGTTGGAACGTGGCGAGAAAT 840
Db 2271 GAATGCGCTGTGATGTCAGTGGCGGCGGATTCACAGCGTTGGAACGTGGCGAGAAAT 2212
QY 841 GAGTTTGGTTGATGAGGAGCTGCGGCAATGCGTGTGCTCCGTGATTTGCTGCCAGCT 900
Db 2211 GAGTTTGGTTGATGAGGAGCTGCGGCAATGCGTGTGCTCCGTGATTTGCTGCCAGCT 2152
QY 901 GGTCTTGGGCGAAGACTTAAGATGCTCAGATGTTTGAAGAGAAATTTGCCCGGCTC 960
Db 2151 GGTCTTGGGCGAAGACTTAAGATGCTCAGATGTTTGAAGAGAAATTTGCCCGGCTC 2092
QY 961 ATGTTGGCGATTCATTCGCGCGCATGCTGATGATGATTTGCACTTGTGGCCTTACGTC 1020
Db 2091 ATGTTGGCGATTCATTCGCGCGCATGCTGATGATGATTTGCACTTGTGGCCTTACGTC 2032
QY 1021 TTGCGGAGATTCCTCGGACTGTCGAGCTGCGGCACTTGGGAAAATCTTCGTGATCAC 1080
Db 2031 TTGCGGAGATTCCTCGGACTGTCGAGCTGCGGCACTTGGGAAAATCTTCGTGATCAC 1972
QY 1081 TTGCATTATGATTTGCCCGCAGCGATGTGCGGCCCCCGATTAATATGCTTCCCAAGTGA 1140
Db 1971 TTGCATTATGATTTGCCCGCAGCGATGTGCGGCCCCCGATTAATATGCTTCCCAAGTGA 1912

RESULT 4

US-10-007-452-5/c
; Sequence 5, Application US/10007452
; Publication No. US20030093701A1
; GENERAL INFORMATION:
; APPLICANT: Tomb, Jean-Francois
; APPLICANT: Bramucci, Michael G.
; APPLICANT: Cheng, Qiong
; APPLICANT: Kostichka, Kristy N.
; TITLE OF INVENTION: Rhodococcus Cloning and Expression Vectors
; FILE REFERENCE: C11709 US NA
; CURRENT APPLICATION NUMBER: US/10/007, 452
; PRIOR FILING DATE: 2001-11-08
; PRIOR APPLICATION NUMBER: 60/254, 868
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 5
; LENGTH: 6334

```
TYPE: DNA
ORGANISM: Rhodococcus AN12
US-10-007-452-5

Query Match
Best Local Similarity 100.0%; Score 1140; DB 14; Length 6334;
Matches 1140; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGACGAGCGTAAGTGTGAACACCTTTCGGCAAGAACCGGCTCCGCTCGTGTGCG 60
DB 3051 ATGACGAGCGTAAGTGTGAACACCTTTCGGCAAGAACCGGCTCCGCTCGTGTGCG 2992
QY 61 TCCGATTAAGCGCGGCAATCCGGGCAAGAACCTTTCGCAATCAACCACTCA 120
DB 2991 TCCGATTAAGCGCGGCAATCCGGGCAAGAACCTTTCGCAATCAACCACTCA 2932
QY 121 GAAACATTTAACCGCTGTGGCCGGGCAATTTCTGGGCGTGAACGGTGTGACATTGTCAAC 180
DB 2931 GAAACATTTAACCGCTGTGGCCGGGCAATTTCTGGGCGTGAACGGTGTGACATTGTCAAC 2872
QY 181 GGTCCGAAAGTTCTGGAATTCGAGGCGCTTCTGCTGCGAAAGGCTGATCTGCCCC 240
DB 2871 GGTCCGAAAGTTCTGGAATTCGAGGCGCTTCTGCTGCGAAAGGCTGATCTGCCCC 2812
QY 241 TGGTGGCGGGAAGTGGGTGACATCGTGCAGAGAAATTTCTGAATGTGTGCTCAT 300
DB 2811 TGGTGGCGGGAAGTGGGTGACATCGTGCAGAGAAATTTCTGAATGTGTGCTCAT 2752
QY 301 CAACCTGGGAGCTGGATCTGTGGATGTGACGATGACCATGCGCATACAGCTGTGAC 360
DB 2751 CAACCTGGGAGCTGGATCTGTGGATGTGACGATGACCATGCGCATACAGCTGTGAC 2692
QY 361 CGGCTCCAGCAGCTTANGACGTGACCTTTGGGCAAGCTGGAAGCTGCCAAGCGTGTG 420
DB 2691 CGGCTCCAGCAGCTTANGACGTGACCTTTGGGCAAGCTGGAAGCTGCCAAGCGTGTG 2632
QY 421 CGTTGGCGGTACGGAAGTGAATGTAACGCTGTGGAGAGATACGTGCGCTGTGTAATC 480
DB 2631 CGTTGGCGGTACGGAAGTGAATGTAACGCTGTGGAGAGATACGTGCGCTGTGTAATC 2572
QY 481 ACTCAGGAAAAAAGCGCTGCGACGTGACGCTTCAAGCTACTCATGTTCACTGATGTGAC 540
DB 2571 ACTCAGGAAAAAAGCGCTGCGACGTGACGCTTCAAGCTACTCATGTTCACTGATGTGAC 2512
QY 541 GTGAGTGAAGAACATCTCGAATCTTCTCGGATGCGATGTGATCGGTGGAATTCGAA 600
DB 2511 GTGAGTGAAGAACATCTCGAATCTTCTCGGATGCGATGTGATCGGTGGAATTCGAA 2452
QY 601 CTGATATCTGTGGGATTTGTGGGCACTACGTAATTTGGGAGTCTCGATGTGACGAAG 660
DB 2451 CTGATATCTGTGGGATTTGTGGGCACTACGTAATTTGGGAGTCTCGATGTGACGAAG 2392
QY 661 ATCGCGGTGAAGCTGATCAAGTTCTCGCTGCGATCTGACGAAATTCATCTGCGTT 720
DB 2391 ATCGCGGTGAAGCTGATCAAGTTCTCGCTGCGATCTGACGAAATTCATCTGCGTT 2332
QY 721 GGTATGAAGTGTGATGAGTGGCGACGGAAGAAAGTGTGCAATGCGAACCGTGTGAC 780
DB 2331 GGTATGAAGTGTGATGAGTGGCGACGGAAGAAAGTGTGCAATGCGAACCGTGTGAC 2272
QY 781 GAAATGCGCTGTGATGTCAGTGGGCGGGGATCCCAAGCGTTGGAATCTGTGGGAGAAATTT 840
DB 2271 GAAATGCGCTGTGATGTCAGTGGGCGGGGATCCCAAGCGTTGGAATCTGTGGGAGAAATTT 2212
QY 841 GAGTTTGTTCGATGAGGAGTGGCGCAATCGCGTGTGCTCGTGTGATGAGTGGCGGAGCT 900
DB 2211 GAGTTTGTTCGATGAGGAGTGGCGCAATCGCGTGTGCTCGTGTGATGAGTGGCGGAGCT 2152
QY 901 GGTCTTTGGGCGAAGCTTAAACAATGCTGATGTTGAGCAGAGAAATCTGCGCGGCTC 960
DB 2151 GGTCTTTGGGCGAAGCTTAAACAATGCTGATGTTGAGCAGAGAAATCTGCGCGGCTC 2092
QY 961 ATGTTTGGATCATTTCCGGCGGAGTGTGATGATGATTCGACTTGTGCGCTTAAGTTC 1020
```

```
DB 2091 ATGTTTGGATCATTTCCGGCGGAGTGTGATGATGATTCGACTTGTGCGCTTAAGTTC 2032
QY 1021 TTCCGCGAGATCTCTGCACTCGTTCGAGCTGGCGGAGCTTGGGAAATCTTGATATCAC 1080
DB 2031 TTCCGCGAGATCTCTGCACTCGTTCGAGCTGGCGGAGCTTGGGAAATCTTGATATCAC 1972
QY 1081 TTGCATTTATGATTCGCGGACGAGATGTGCGGCCCGGATATATATCGTTTGCAGTGA 1140
DB 1971 TTGCATTTATGATTCGCGGACGAGATGTGCGGCCCGGATATATATCGTTTGCAGTGA 1912

RESULT 5
US-10-007-527a-7/c
Sequence 7, Application US/10007527A
Publication No. US20030044807A1
GENERAL INFORMATION:
APPLICANT: Tomb, Jean-Francois
APPLICANT: Bramucci, Michael G.
APPLICANT: Cheng, Qilong
APPLICANT: Koshchka, Kristy N.
TITLE OF INVENTION: Rhodococcus Cloning and Expression Vectors
FILE REFERENCE: C14709 US NA
CURRENT APPLICATION NUMBER: US/10/007,527A
PRIOR FILING DATE: 2001-12-05
PRIOR APPLICATION NUMBER: 60/254,868
PRIOR FILING DATE: 2000-12-12
NUMBER OF SEQ ID NOS: 30
SOFTWARE: Microsoft Office 97
SEQ ID NO 7
LENGTH: 9652
TYPE: DNA
ORGANISM: Plasmid pRHR17
US-10-007-527a-7

Query Match
Best Local Similarity 100.0%; Score 1140; DB 14; Length 9652;
Matches 1140; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGACGAGCGTAAGTGTGAACACCTTTCGGCAAGAACCGGCTCCGCTCGTGTGCG 60
DB 5051 ATGACGAGCGTAAGTGTGAACACCTTTCGGCAAGAACCGGCTCCGCTCGTGTGCG 4993
QY 61 TCCGATTAAGCGCGGCAATCCGGGCAAGAACCTTTCGCAATCAACCACTCA 120
DB 4992 TCCGATTAAGCGCGGCAATCCGGGCAAGAACCTTTCGCAATCAACCACTCA 4933
QY 121 GAAACATTTAACCGCTGTGGCCGGGCAATTTCTGGGCGTGAACGGTGTGACATTGTCAAC 180
DB 4932 GAAACATTTAACCGCTGTGGCCGGGCAATTTCTGGGCGTGAACGGTGTGACATTGTCAAC 4873
QY 181 GGTCCGAAAGTGTGGAATTCGAGGCGCTTCTGCTGCGAAAGGCTGATCTGCCCC 240
DB 4872 GGTCCGAAAGTGTGGAATTCGAGGCGCTTCTGCTGCGAAAGGCTGATCTGCCCC 4813
QY 241 TGGTGGCGGGAAGTGGGTGACATCGTGCAGAGAAATTTCTGAATGTGTGCTCAT 300
DB 4812 TGGTGGCGGGAAGTGGGTGACATCGTGCAGAGAAATTTCTGAATGTGTGCTCAT 4753
QY 301 CAACCTGGGAGCTGGATCTGTGGATGTGACGATGACCATGCGCATACAGCTGTGAC 360
DB 4752 CAACCTGGGAGCTGGATCTGTGGATGTGACGATGACCATGCGCATACAGCTGTGAC 4693
QY 361 CGGCTCCAGCAGCTTANGACGTGACCTTTGGGCAAGCTGGAAGCTGCCAAGCGTGTG 420
DB 4692 CGGCTCCAGCAGCTTANGACGTGACCTTTGGGCAAGCTGGAAGCTGCCAAGCGTGTG 4633
QY 421 CGTTGGCGGTACGGAAGTGAATGTAACGCTGTGGAGAGATACGTGCGCTGTGTAATC 480
DB 4632 CGTTGGCGGTACGGAAGTGAATGTAACGCTGTGGAGAGATACGTGCGCTGTGTAATC 4573
QY 481 ACTCAGGAAAAAAGCGCTGCGACGTGACGCTTCAAGCTACTCATGTTGATGTGAC 540
```

Db 4572 ACTCAGGAAAAACGGCTGCGACGTCCACGTTCAAGCGCTACTCATGTTTCAAGTGTGAC 4513
QY 541 GTGAGTGAGACATCTCTGCAATCTCTCTGGAATGCAATGTTGATGCGTGGACTTCCAA 600
Db 4512 GTGAGTGAGACATCTCTGCAATCTCTCTGGAATGCAATGTTGATGCGTGGACTTCCAA 4453
QY 601 CTCGTATCTCTGGAATTTGCTGCGCACTACAGTAATTCGGATGCTGATGTAAGAAAG 660
Db 4452 CTCGTATCTCTGGAATTTGCTGCGCACTACAGTAATTCGGATGCTGATGTAAGAAAG 4393
QY 661 ATCGGCGGTGAAGCTGATCAAGTCTCGTGCATCTGACGAAATTCATCTGCGCTT 720
Db 4392 ATCGGCGGTGAAGCTGATCAAGTCTCGTGCATCTGACGAAATTCATCTGCGCTT 4333
QY 721 GGTATGAGAGTTGTTGTTGCGGCGACGAAAAAGTGTGACATGCGCAACCGTGCACCTG 780
Db 4332 GGTATGAGAGTTGTTGTTGCGGCGACGAAAAAGTGTGACATGCGCAACCGTGCACCTG 4273
QY 781 GAAATCGCTGTGATGATGCGGCGGCGGATCCACAAAGCGTTGAACTGTGCGGAAATTT 840
Db 4272 GAAATCGCTGTGATGATGCGGCGGCGGATCCACAAAGCGTTGAACTGTGCGGAAATTT 4213
QY 841 GAGTTGTTGATGATGCGGCGGCGGATCCACAAAGCGTTGAACTGTGCGGAAATTT 900
Db 4212 GAGTTGTTGATGATGCGGCGGCGGATCCACAAAGCGTTGAACTGTGCGGAAATTT 4153
QY 901 GGTCTTTGGGCGAACTTAACAGATGCTGATGTTGAGCAGAAAGATCTGCGCGCTC 960
Db 4152 GGTCTTTGGGCGAACTTAACAGATGCTGATGTTGAGCAGAAAGATCTGCGCGCTC 4093
QY 961 ATGTTGCGATCATTCGCGCGCGGCGGATCGTGGATGATGATTCGAGCTTGAGCGGCTTAACGTC 1020
Db 4092 ATGTTGCGATCATTCGCGCGCGGCGGATCGTGGATGATGATTCGAGCTTGAGCGGCTTAACGTC 4033
QY 1021 TTCGCGAGATCTCTGCACTGCTGCAAGCTGCGCGCATTTGGAAAAATCTTGATGATC 1080
Db 4032 TTCGCGAGATCTCTGCACTGCTGCAAGCTGCGCGCATTTGGAAAAATCTTGATGATC 3973
QY 1081 TTGCATTATCGATTGCGCGCGGCGGATGTCGCGCGCGGATTAATCGGTTGCAAGTGA 1140
Db 3972 TTGCATTATCGATTGCGCGCGGCGGATGTCGCGCGCGGATTAATCGGTTGCAAGTGA 3913

RESULT 6
US-10-007-452-7/c
; Sequence 7, Application US/10007452
; Publication No. US20030093701A1
; GENERAL INFORMATION:
; APPLICANT: Tomb, Jean-Francois
; APPLICANT: Bramucci, Michael G.
; APPLICANT: Cheng, Qilong
; APPLICANT: Koestlichka, Kristy N.
; TITLE OF INVENTION: Rhodococcus Cloning and Expression Vectors
; FILE REFERENCE: CL1709 US NA
; CURRENT APPLICATION NUMBER: US/10/007,452
; PRIOR APPLICATION NUMBER: 60/254,868
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 7
; LENGTH: 9652
; TYPE: DNA
; ORGANISM: P1aamid prHR17
US-10-007-452-7

Query Match 100.0%; Score 1140; DB 14; Length 9652;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1140; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGACGAGCTGATGCTGACACCTTCCGCGAAAGACGCGCTCCGCTCGTGTG 60
Db 5052 ATGACGAGCTGATGCTGACACCTTCCGCGAAAGACGCGCTCCGCTCGTGTG 4993

QY 61 TCCGATTAACCGGCGATCCGCGACGAACTGCGAACCCAACTTCAACAAATCAACAGTCA 120
Db 4992 TCCGATTAACCGGCGATCCGCGACGAACTGCGAACCCAACTTCAACAAATCAACAGTCA 4933
QY 121 GAAACATTTAAGCGCTGCGCGCGGATTTCTGCGGTAAGCGGTGTGACATTTGTCAAC 180
Db 4932 GAAACATTTAAGCGCTGCGCGCGGATTTCTGCGGTAAGCGGTGTGACATTTGTCAAC 4873
QY 181 GGTCCGAAAGGTTCTGATTCGAGGCGCTTGCTTCTGCGGAAAGGCTGATCTGCGCC 240
Db 4872 GGTCCGAAAGGTTCTGATTCGAGGCGCTTGCTTCTGCGGAAAGGCTGATCTGCGCC 4813
QY 241 TCGTGTGCGGAAAAAGTGTGCACTGCGTGAACGAAATTTCTCAAGTTGTTGCTCAT 300
Db 4812 TCGTGTGCGGAAAAAGTGTGCACTGCGTGAACGAAATTTCTCAAGTTGTTGCTCAT 4753
QY 301 CAACCTGCGGACTGATCTGTTGCGATGCTGACATGACCATGCGCATACAGCTGTGAC 360
Db 4752 CAACCTGCGGACTGATCTGTTGCGATGCTGACATGACCATGCGCATACAGCTGTGAC 4693
QY 361 CGGCTCCAGACCTATGAGCTGACCTTCGCGAGCGCTGGAAGCTGCGACCAACGCTGT 420
Db 4692 CGGCTCCAGACCTATGAGCTGACCTTCGCGAGCGCTGGAAGCTGCGACCAACGCTGT 4633
QY 421 CGTGTGCTACGGAACGTGAATGTACGCTGCGACGATACGTCGCGCTGTTGAAATC 480
Db 4632 CGTGTGCTACGGAACGTGAATGTACGCTGCGACGATACGTCGCGCTGTTGAAATC 4573
QY 481 ACTGACGGAATAAACGCTGCGACGTCACGCTTCAACGCTCATATGTTCAAGTGTAC 540
Db 4572 ACTGACGGAATAAACGCTGCGACGTCACGCTTCAACGCTCATATGTTCAAGTGTAC 4513
QY 541 GTGAGTGAGACATCTCTGCAATCTCTCGATGCGATGCGATGCGTGGACTTCCAA 600
Db 4512 GTGAGTGAGACATCTCTGCAATCTCTCGATGCGATGCGATGCGTGGACTTCCAA 4453
QY 601 CTCGTATCTCTGGAATTTGCTGCGCACTACAGTAATTCGGATGCTGATGTAAGAAAG 660
Db 4452 CTCGTATCTCTGGAATTTGCTGCGCACTACAGTAATTCGGATGCTGATGTAAGAAAG 4393
QY 661 ATCGGCGGTGAAGCTGATCAAGTCTCGTGCATCTGACGAAATTCATCTGCGCTT 720
Db 4392 ATCGGCGGTGAAGCTGATCAAGTCTCGTGCATCTGACGAAATTCATCTGCGCTT 4333
QY 721 GGTATGAGAGTTGTTGTTGCGGCGACGAAAAAGTGTGACATGCGCAACCGTGCACCTG 780
Db 4332 GGTATGAGAGTTGTTGTTGCGGCGACGAAAAAGTGTGACATGCGCAACCGTGCACCTG 4273
QY 781 GAAATCGCTGTGATGATGCGGCGGCGGATCCACAAAGCGTTGAACTGTGCGGAAATTT 840
Db 4272 GAAATCGCTGTGATGATGCGGCGGCGGATCCACAAAGCGTTGAACTGTGCGGAAATTT 4213
QY 841 GAGTTGTTGATGATGCGGCGGCGGATGTCGCGCGGATTAATCGGTTGCAAGTGA 900
Db 4212 GAGTTGTTGATGATGCGGCGGCGGATGTCGCGCGGATTAATCGGTTGCAAGTGA 4153
QY 901 GGTCTTTGGGCGAACTTAACAGATGCTGATGTTGAGCAGAAAGATCTGCGCGCTC 960
Db 4152 GGTCTTTGGGCGAACTTAACAGATGCTGATGTTGAGCAGAAAGATCTGCGCGCTC 4093
QY 961 ATGTTGCGATCATTCGCGCGCGGCGGATCGTGGATGATGATTCGAGCTTGAGCGGCTTAACGTC 1020
Db 4092 ATGTTGCGATCATTCGCGCGCGGCGGATCGTGGATGATGATTCGAGCTTGAGCGGCTTAACGTC 4033
QY 1021 TTCGCGAGATCTCTGCACTGCTGCAAGCTGCGCGCATTTGGAAAAATCTTGATGATC 1080
Db 4032 TTCGCGAGATCTCTGCACTGCTGCAAGCTGCGCGCATTTGGAAAAATCTTGATGATC 3973
QY 1081 TTGCATTATCGATTGCGCGCGGCGGATGTCGCGCGGATTAATCGGTTGCAAGTGA 1140
Db 3972 TTGCATTATCGATTGCGCGCGGCGGATGTCGCGCGGATTAATCGGTTGCAAGTGA 3913

RESULT 7
US-10-007-527A-6/c
; Sequence 6, Application US/10007527A
; Publication No. US20030044807A1
; GENERAL INFORMATION:
; APPLICANT: Tomb, Jean-Francois
; APPLICANT: Bramucci, Michael G.
; APPLICANT: Cheng, Qiong
; APPLICANT: Kostichka, Kristy N.
; TITLE OF INVENTION: Rhodococcus Cloning and Expression Vectors
; FILE REFERENCE: CL1709 US NA
; CURRENT APPLICATION NUMBER: US/10/007,527A
; PRIOR FILING DATE: 2001-12-05
; PRIOR APPLICATION NUMBER: 60/254,868
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 6
; LENGTH: 11241
; TYPE: DNA
; ORGANISM: Plasmid pRHB17
US-10-007-527A-6

Query Match 100.0%; Score 1140; DB 14; Length 11241;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1140; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGACGAGGTAAAGTGTGACACCTTCCGGCAAAAGCCGCGCTCCGCTCGTGTG 60
DB 6641 ATGACGAGGTAAAGTGTGACACCTTCCGGCAAAAGCCGCGCTCCGCTCGTGTG 6582
QY 61 TCCGATTAAGCGCGGATCCGCGACGAACTGCGACCCAACTTCAACAAATCAACAAGTCA 120
DB 6581 TCCGATTAAGCGCGGATCCGCGACGAACTGCGACCCAACTTCAACAAATCAACAAGTCA 6522
QY 121 GAAACATTTAAGCCTGTGTGCGCGCGCGATTTCTGCGGTGAACGCTGTGAACATTGTCAAC 180
DB 6521 GAAACATTTAAGCCTGTGTGCGCGCGCGATTTCTGCGGTGAACGCTGTGAACATTGTCAAC 6462
QY 181 GGTCCGAAAGTTCGTGATTCGGAAGCCCTTCGTTCTCGGGGAAAGGGCTGTGATTCGCCCC 240
DB 6461 GGTCCGAAAGTTCGTGATTCGGAAGCCCTTCGTTCTCGGGGAAAGGGCTGTGATTCGCCCC 6402
QY 241 TCGTGTGCGGAAAGTTCGTGATTCGGAAGCCCTTCGTTCTCGGGGAAAGGGCTGTGATTCGCCCC 300
DB 6401 TCGTGTGCGGAAAGTTCGTGATTCGGAAGCCCTTCGTTCTCGGGGAAAGGGCTGTGATTCGCCCC 6342
QY 301 CAACCTCGGAGTGTGATTCGTGATTCGGAAGCCCTTCGTTCTCGGGGAAAGGGCTGTGATTCGCCCC 360
DB 6341 CAACCTCGGAGTGTGATTCGTGATTCGGAAGCCCTTCGTTCTCGGGGAAAGGGCTGTGATTCGCCCC 6282
QY 361 CGGCTCCAGACCTATGGAATCTTCCGCGAGCCTTGAAGAGCTGCAACCGTCCGT 420
DB 6281 CGGCTCCAGACCTATGGAATCTTCCGCGAGCCTTGAAGAGCTGCAACCGTCCGT 6222
QY 421 CGTGGCGTACGGAAGTGTGAATGTACGCGCTGCGACGGAATACGTCGCGCTGTGAATC 480
DB 6221 CGTGGCGTACGGAAGTGTGAATGTACGCGCTGCGACGGAATACGTCGCGCTGTGAATC 6162
QY 481 ACTCAGGAAAGAAAGCGGCTGCGACGTCACGTTCAAGCGCTACTCATGTTCAGTGTGAC 540
DB 6161 ACTCAGGAAAGAAAGCGGCTGCGACGTCACGTTCAAGCGCTACTCATGTTCAGTGTGAC 6102
QY 541 GTGAGTGAAGACATCTCTCGAATTCCTTCCGAGTGCATGTTCATGCGTGTGACAA 600
DB 6101 GTGAGTGAAGACATCTCTCGAATTCCTTCCGAGTGCATGTTCATGCGTGTGACAA 6042
QY 601 CTGATCTCTGGGATTTGCTGCGCACTACGTAATTCGCGTGTGATTCGATGTGACAAAG 660
DB 6041 CTGATCTCTGGGATTTGCTGCGCACTACGTAATTCGCGTGTGATTCGATGTGACAAAG 5982
QY 661 ATCGCGGTGAAGCTGATCAAGTTCTCGCTGCGTATCTGACGAAATTCGATTCGCGCTT 720

DB 5981 ATCGCGGTGAAGCTGATCAAGTTCTCGCTGCGTATCTGACGAAATTCGATTCGCGCTT 5922
QY 721 GGTATGAGGTTGGTATGCGGACGGAAGAAAGGTGTGACATTCGCAACCGTGCACCTCG 780
DB 5921 GGTATGAGGTTGGTATGCGGACGGAAGAAAGGTGTGACATTCGCAACCGTGCACCTCG 5862
QY 781 GAAATCGTGTGATGACAGTGGCGGGATCCAGAGCGTGTGAATCTGTGCGAGATTT 840
DB 5861 GAAATCGTGTGATGACAGTGGCGGGATCCAGAGCGTGTGAATCTGTGCGAGATTT 5802
QY 841 GAGTTGGTGTGATGAGACGTGCGCAATTCGCGTGTGATTCGCGTGTGACCGGAGCT 900
DB 5801 GAGTTGGTGTGATGAGACGTGCGCAATTCGCGTGTGATTCGCGTGTGACCGGAGCT 5742
QY 901 GGTCTGGGGGAGAACTAAGATGCTGAGATTCGTTGAGAGGAAATTCGCGCGCTC 960
DB 5741 GGTCTGGGGGAGAACTAAGATGCTGAGATTCGTTGAGAGGAAATTCGCGCGCTC 5682
QY 961 ATGATTGCGATCATTCGCGCGGATTCGTTGAGATTCGTTGAGATTCGTTGAGATTC 1020
DB 5681 ATGATTGCGATCATTCGCGCGGATTCGTTGAGATTCGTTGAGATTCGTTGAGATTC 5622
QY 1021 TTGCGCGAGATTCCTGGAACCTGTCGAACTGCGCGGATTCGTTGAGATTCGTTGAGATTC 1080
DB 5621 TTGCGCGAGATTCCTGGAACCTGTCGAACTGCGCGGATTCGTTGAGATTCGTTGAGATTC 5562
QY 1081 TTGCAATTCGATTCGCGCGGATTCGTTGAGATTCGTTGAGATTCGTTGAGATTC 1140
DB 5561 TTGCAATTCGATTCGCGCGGATTCGTTGAGATTCGTTGAGATTCGTTGAGATTC 5502

RESULT 8
US-10-007-452-6/c
; Sequence 6, Application US/10007452
; Publication No. US20030093701A1
; GENERAL INFORMATION:
; APPLICANT: Tomb, Jean-Francois
; APPLICANT: Bramucci, Michael G.
; APPLICANT: Cheng, Qiong
; APPLICANT: Kostichka, Kristy N.
; TITLE OF INVENTION: Rhodococcus Cloning and Expression Vectors
; FILE REFERENCE: CL1709 US NA
; CURRENT APPLICATION NUMBER: US/10/007,452
; PRIOR FILING DATE: 2001-11-08
; PRIOR APPLICATION NUMBER: 60/254,868
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 6
; LENGTH: 11241
; TYPE: DNA
; ORGANISM: Plasmid pRHB17
US-10-007-452-6

Query Match 100.0%; Score 1140; DB 14; Length 11241;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1140; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGACGAGGTAAAGTGTGACACCTTCCGGCAAAAGCCGCGCTCCGCTCGTGTG 60
DB 6641 ATGACGAGGTAAAGTGTGACACCTTCCGGCAAAAGCCGCGCTCCGCTCGTGTG 6582
QY 61 TCCGATTAAGCGCGGATCCGCGACGAACTGCGACCCAACTTCAACAAATCAACAAGTCA 120
DB 6581 TCCGATTAAGCGCGGATCCGCGACGAACTGCGACCCAACTTCAACAAATCAACAAGTCA 6522
QY 121 GAAACATTTAAGCCTGTGTGCGCGCGCGATTTCTGCGGTGAACGCTGTGAACATTGTCAAC 180
DB 6521 GAAACATTTAAGCCTGTGTGCGCGCGCGATTTCTGCGGTGAACGCTGTGAACATTGTCAAC 6462
QY 181 GGTCCGAAAGTTCGTGATTCGGAAGCCCTTCGTTCTCGGGGAAAGGGCTGTGATTCGCCCC 240

```
Db 6461 GGTCCGAAAGGTTCTGGATTCCGAGGCTTTCGTTCTCGGAGAAAGGCTGATCTGCCCC 6402
Qy 241 TGGTGGCGGAGAAAGCGGTGCAATCGTGGAGAGAAATTTCTCAAGTTGTGCTCAT 300
Db 6401 TGGTGGCGGAGAAAGCGGTGCAATCGTGGAGAGAAATTTCTCAAGTTGTGCTCAT 6342
Qy 301 CCACTCGGAGATCGATCTGTGGATGGTGAACATGAGCCATCAGTGGTCCAG 360
Db 6341 CCACTCGGAGATCGATCTGTGGATGGTGAACATGAGCCATCAGTGGTCCAG 6282
Qy 361 CGGCTCCACGACCTTATGATGATGACCTTTCGCGAGCTTGGAAAGCTCGACCAAGCTGCT 420
Db 6281 CGGCTCCACGACCTTATGATGATGACCTTTCGCGAGCTTGGAAAGCTCGACCAAGCTGCT 6222
Qy 421 CGTTGGCGTACGGAAGGTGAATGTACGGCTGCGACGATACGTCGCGCTGTGAATC 480
Db 6221 CGTTGGCGTACGGAAGGTGAATGTACGGCTGCGACGATACGTCGCGCTGTGAATC 6162
Qy 481 ACTCAGGAGAAAGGCTGCGACGCTGCGACGCTTCAAGGCTCTCATGTTCAAGTGTGAC 540
Db 6161 ACTCAGGAGAAAGGCTGCGACGCTGCGACGCTTCAAGGCTCTCATGTTCAAGTGTGAC 6102
Qy 541 GTGAGTGAAGAACCTCTCGAATCTTCTCGATGCGATGTTGATCGATCGTGAAGCTTCAAA 600
Db 6101 GTGAGTGAAGAACCTCTCGAATCTTCTCGATGCGATGTTGATCGATCGTGAAGCTTCAAA 6042
Qy 601 CTCGATCTCTGGAGATTTGCTGCGACCTAAGTAACTGAGGTGCTCGATGACGAAG 660
Db 6041 CTCGATCTCTGGAGATTTGCTGCGACCTAAGTAACTGAGGTGCTCGATGACGAAG 5982
Qy 661 ATCGGCGGTGAAGCTGATCAAGTTCCTCGCTGCGATCTGACGAAATTTGATCTGCGCTT 720
Db 5981 ATCGGCGGTGAAGCTGATCAAGTTCCTCGCTGCGATCTGACGAAATTTGATCTGCGCTT 5922
Qy 721 GGTATGAGAGTTGTGATGAGCGACGAGAAAGTGTGCAATGACGACCGTCAAGCTTGG 780
Db 5921 GGTATGAGAGTTGTGATGAGCGACGAGAAAGTGTGCAATGACGACCGTCAAGCTTGG 5862
Qy 781 GAAATCGCTGTTGATGATGAGCGCGGAGATCCAAAGCTTGGAACTGTGCGAGAAATTT 840
Db 5861 GAAATCGCTGTTGATGATGAGCGCGGAGATCCAAAGCTTGGAACTGTGCGAGAAATTT 5802
Qy 841 GAGTTGGTTCGATGAGGACGTGCGGCAATCGGCTGCTCCCGGATGTTGCTGCGAGCT 900
Db 5801 GAGTTGGTTCGATGAGGACGTGCGGCAATCGGCTGCTCCCGGATGTTGCTGCGAGCT 5742
Qy 901 GGTCTTGGGCGAGAACTTAACAGATGCTCAGATCGTTGAGCAGAAATCTCCCGGCTC 960
Db 5741 GGTCTTGGGCGAGAACTTAACAGATGCTCAGATCGTTGAGCAGAAATCTCCCGGCTC 5682
Qy 961 ATGTTGGCATCATTCGCGCGGATCGTGGATGATGATTCGAGCTTGTGCGCTTACGTC 1020
Db 5681 ATGTTGGCATCATTCGCGCGGATCGTGGATGATGATTCGAGCTTGTGCGCTTACGTC 5622
Qy 1021 TTCGCGAGATCCTCGGACTCGTCAAGCTGCGCGAGACTTGGGAAATCTTGTGATCAC 1080
Db 5621 TTCGCGAGATCCTCGGACTCGTCAAGCTGCGCGAGACTTGGGAAATCTTGTGATCAC 5562
Qy 1081 TTGCATTATCGATTGCGCGAGCGAGTGGCGGCGCGAGTAAATCGTTGCAAGTGA 1140
Db 5561 TTGCATTATCGATTGCGCGAGCGAGTGGCGGCGCGAGTAAATCGTTGCAAGTGA 5502

RESULT 9
US-10-184-644-332
; Sequence 332, Application US/10184644
; Publication No. US20030044930A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jjian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Godowski, Paul J.
```

```
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3430R1C227
; CURRENT APPLICATION NUMBER: US/10/184,644
; PRIOR APPLICATION REMOVED - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 332
; LENGTH: 520
; TYPE: PR
; ORGANISM: Homo Sapien
US-10-184-644-332
```

```
Query Match 3.6%; Score 40.8; DB 14; Length 520;
Best Local Similarity 18.5%; Pred. No. 0.0098;
Matches 74; Conservative 81; Mismatches 246; Indels 0; Gaps 0;
```

```
Qy 522 ACTCATGTTCAAGTGTGACGTGAGTGAACAATCTCGAATCTTCTCGATGCGATGTT 581
Db 90 DSSPSLLQSAHPGSHLQGASRLQVLQALTWVAVSHHLLQRYDNFTQNPMPRIKGB 149
Qy 582 CGATCGTGAAGTCTCAACTGCTGATCTCTGGATTTGCTGGCCACTACGTAATTCGG 641
Db 150 QGAPRLQGHKGMKMPGAPRPPGPAKAKAMRGDAGTGSPPGPPVGEAGLQGP 209
Qy 642 TGGTCTCGATGATGACAAAGATCGCGGTGAAGCTGATCAAGTTCGCTGCGTATCTGAC 701
Db 210 QGAPRQKQATGTPPQGGKSGKDGGLIGPKETGTGKBEKDLGLPGSGDKGMKGDAGV 269
Qy 702 GAAATTCATCTGCGCTGTTGATGAGAGTGTGTGTGCGCAGCGAAAGTGTGCGACA 761
Db 270 MGPFGAGSKADPFGPPGPPGAGPFGAGDQGPGLQGPFPFGAVGHPGAKGEPGSGS 329
Qy 762 TGGCAACGCTGCAACCTCGGAAATGCTGTTGATGACAGGCGCGGAGATCCACAGCGTT 821
Db 330 PORAGLPSPSPGPAATGLKSGKBDGLQDQGRKBSGVPFGAVGSGSPGLAPKGA 389
Qy 822 GGAATGTGGCGAGAAATTTGAATTTGATTCGATGAGAGCTCGGCAATCGCTGTCGCG 881
Db 390 PQAAGQKGDQGVKSSSGHQGVKGEKHEGEMSVSVRIYSSNRGAALFYYSGTWGTICDD 449
Qy 882 TGAATTCGCTGCGCGAGCTGCTTGTGGGCGAGAACTAACAG 922
Db 450 EMONSDAIVFCRMIGYSKGRALYKVAAGTGOIWLNDNVQCRG 490
```

```
RESULT 10
US-10-184-634-332
; Sequence 332, Application US/10184634
; Publication No. US20030068684A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jjian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3430R1C217
; CURRENT APPLICATION NUMBER: US/10/184,634
; CURRENT FILING DATE: 2002-06-28
```

;; Prior Application removed - See File Wrapper or Palm
;; NUMBER OF SEQ ID NOS: 612
;; SEQ ID NO 332
;; LENGTH: 520
;; TYPE: PRT
;; ORGANISM: Homo Sapien
US-10-184-634-332

Query Match 3.6%; Score 40.8; DB 14; Length 520;
Best Local Similarity 18.5%; Pred. No. 0.0098;
Matches 74; Conservative 81; Mismatches 246; Indels 0; Gaps 0;

QY 522 ACTCATGTTCAAGTGTGACGTGAGTGAACAATCCTCGAATCCTTCTCGATGCGATGTT 581
DB 90 DSSPSFLQSAHPBHEHLAQASRLQYLQQLTWVRVSHHLLQRYDNFQNGMRKIBE 149
QY 582 CGATCGTGGACTTCCAAACTGCTGATCTCTGGATTGCTGGCCACTACGTAATTCGGG 641
DB 150 QGAPGLQGHKGMGMGAPGPPGPPAEXKAGKAMGRDGTGSPGQSPGVKEXAGLQGP 209
QY 642 TGCTCGATGATGACGAAGAATCGGCGGTGAAGCTGATCAAGTTCTGCTGCGATCTGAC 701
DB 210 QGAPGKQATGTPGPGEGESKDGGLIPKGETGKGRKDLGLPESKDRGMKDDAGV 269
QY 702 GAAATTTGATCTGGCGTTGATATGAGAGTTGATGAGCGAAGAAAAGTGTGACA 761
DB 270 MGPAGQSKKDPGRGPPGLAGFPAGKDDQGPGLQVPPGAVGHGAGBPSAGS 329
QY 762 TGGCAACCTGCACCTCGGAAATCGCTTGTATGAGAGTGGCGGAGATCCACACCTT 821
DB 330 PPARGLPGSPSGATGLKSKDGTGLQOQGRKSSVPPGPGVGEQSSPGLAPKGA 389
QY 822 GGAACGTGGCGGAATTTGAGTTGATTGATGAGACGTGGGCAATCGCTGCTCCG 881
DB 390 PQAGGKGPQGVKSSGEGSVKKEKERGENSVIVSSNRGRBYVYSGTWGTCDD 449
QY 882 TGGATTGCGTCCGAGCTGCTTGGGCGAAGACTACAG 922
DB 450 EMONSDAIVFCRMLGYSKGRALYKVYAGTGQIWLDMVQCRG 490

RESULT 11

US-10-140-472-10
;; Sequence 10, Application US/10140472
;; Publication No. US20030138888A1
;; GENERAL INFORMATION:
;; APPLICANT: Baker, Kevin P.
;; APPLICANT: Beresini, Maureen
;; APPLICANT: DeForge, Laura
;; APPLICANT: Desnoyers, Luc
;; APPLICANT: Filvaroff, Ellen
;; APPLICANT: Gao, Wei-Qiang
;; APPLICANT: Gerritsen, Mary E.
;; APPLICANT: Goddard, Audrey
;; APPLICANT: Godowski, Paul J.
;; APPLICANT: Gurney, Austin L.
;; APPLICANT: Sherwood, Steven
;; APPLICANT: Smith, Victoria
;; APPLICANT: Stewart, Timothy A.
;; APPLICANT: Tumas, Daniel
;; APPLICANT: Watanabe, Colin K
;; APPLICANT: Wood, William
;; APPLICANT: Zhang, Zemin
;; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
;; TITLE OF INVENTION: ACIDS ENCODING THE SAME
;; FILE REFERENCE: P3330R1C168
;; CURRENT APPLICATION NUMBER: US/10/140,472
;; CURRENT FILING DATE: 2002-05-06
;; Prior Application removed - See File Wrapper or Palm
;; NUMBER OF SEQ ID NOS: 550
;; SEQ ID NO 10
;; LENGTH: 594
;; TYPE: PRT

;; ORGANISM: Homo Sapien
US-10-140-472-10

Query Match 3.3%; Score 37.6; DB 12; Length 594;
Best Local Similarity 9.4%; Pred. No. 0.12;
Matches 36; Conservative 145; Mismatches 204; Indels 0; Gaps 0;

QY 692 CGTATCTGACGAAATTCATCTGCGTTGATGAGAGTGTGATGCGACGAAATAA 751
DB 195 SDBEIVIPDIDVVDVBLNQEYADLNKQATITGMADGDVRLNRKDEAEAIKAKA 254
QY 752 GTGTGACATGCGACCGTGCACCTGGAAATCGCTGTGATGACAGTGGCGGGATC 811
DB 255 LEEBKAMYSGRHSRQRREFREKRLRGKISPPSYARADSPYDPYKRSPESSSSRSR 314
QY 812 CACAAAGCTTGAATCTGCGGAGAAATTTGATTTGTTGATGATGAGAGTGGCGAAATCG 871
DB 315 SRSPTPGEEKITFTFTSGSDDEAAAAAAGVTTGPPAPDPGPAPGNASAR 374
QY 872 CGTGTCCTCGTATGATGCGTCCGAGCTGCTTGGGCGACAACTACAGATGCTCAGA 931
DB 375 RRSSSSSSSASRTSSSSSSSSSSRRRGGYRSGRHASRSRSMRSRBRYSR 434
QY 932 TCGTTGACAGAGAAATCTGCCCGCTCATGTTGATTCGATCTCCGCGCATGTGA 991
DB 435 SRSRGRHRSRSGSRDGHYRSRPARGGYGPRRSRSSRSHSGDRYRGGRLRHSSRS 494
QY 992 TGATGATTCGAGCTTGTGCGGCTTACGTTCTCGGCGAATCTCGCATCTGTCGAAGTCG 1051
DB 495 RSMWLSFRSRSLTRSHSPSPGSRSSRSRSPSPAPARKLTPPAAPVAGEL 554
QY 1052 GCGCAGCTTGGAAATCTTCGTGA 1076
DB 555 KTEBPAGKETGAAKVTQADASGEA 579

RESULT 12

US-10-141-761-10
;; Sequence 10, Application US/10141761
;; Publication No. US20030148432A1
;; GENERAL INFORMATION:
;; APPLICANT: Baker, Kevin P.
;; APPLICANT: Beresini, Maureen
;; APPLICANT: DeForge, Laura
;; APPLICANT: Desnoyers, Luc
;; APPLICANT: Filvaroff, Ellen
;; APPLICANT: Gao, Wei-Qiang
;; APPLICANT: Gerritsen, Mary E.
;; APPLICANT: Goddard, Audrey
;; APPLICANT: Godowski, Paul J.
;; APPLICANT: Gurney, Austin L.
;; APPLICANT: Sherwood, Steven
;; APPLICANT: Smith, Victoria
;; APPLICANT: Stewart, Timothy A.
;; APPLICANT: Tumas, Daniel
;; APPLICANT: Watanabe, Colin K
;; APPLICANT: Wood, William
;; APPLICANT: Zhang, Zemin
;; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
;; TITLE OF INVENTION: ACIDS ENCODING THE SAME
;; FILE REFERENCE: P3330R1C198
;; CURRENT APPLICATION NUMBER: US/10/141,761
;; CURRENT FILING DATE: 2002-05-08
;; Prior Application removed - See Palm or File Wrapper
;; NUMBER OF SEQ ID NOS: 550
;; SEQ ID NO 10
;; LENGTH: 594
;; TYPE: PRT
;; ORGANISM: Homo Sapien
US-10-141-761-10

Query Match 3.3%; Score 37.6; DB 12; Length 594;
Best Local Similarity 9.4%; Pred. No. 0.12;

```

QY      752 GTGGTCGACATGGCGAACCGTGACACCTGGGAAATCGCTGTGATGACGCGCGGGATC 811
Db      255 LEEBKAMYSGRRSRQRGRFRERKRLRKRTISPBYARDSPITYPYKRSBESSSESRSR 314
QY      812 CACAGCGTTGAACTGTGCGGAGATTTGATGTTGTTGCATGAGGACGTGCGGCAATCG 871
Db      315 SRSPTPGREKRTFTTSPFGGSDDEFAAAAAAAAGVTTGKPPAPPGGAPGRNMSAR 374
QY      872 CGTGTCCCGTGCATTTCCGTGCCGAGCTGTCTTGCGGCGAGAACTACAGATCTCAGA 931
Db      375 RRSSTSSSSSSASATSSRSRSSRSRSGGAYRRGRHARSRSRWSRSRSRYSR 434
QY      932 TCGTTGACGAGAAATATCGCCCCCGTCATCGTTGGCATATCCGCGCGAGATCGGA 991
Db      435 SRSGRRHSGGSGSDGHRYSRSPARRGGYPPRRSRSSRSBSGDRYRRGRGLRHHSSRS 494
QY      992 TGATGATTCGACACTTGTGCGCCTTACGTCTTCCGCGAGATCTTCGGACTCGTCAGACTG 1051
Db      495 RSSWSLSPSRSRSTRTSRSHSPSPSGRSRSRSGSPSPSPAREKLTLPAPASRAVERKL 554
QY      1052 GCGGCACTTGGGAAATCTTCGTGA 1076
Db      555 KTEPPAGKETGAAKVTQADASGBA 579

RESULT 14
US-10-158-790-10
; Sequence 10, Application US/10158790
; Publication No. US20030180879A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Flaviaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P330R1C448
; CURRENT APPLICATION NUMBER: US/10/158,790
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 10
; LENGTH: 594
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-158-790-10

Query Match      3 3%, Score 37.6, DB 12, Length 594;
Best Local Similarity 9.4%, Pred.No.0.12;
Matches 36; Conservative 145; Mismatches 204; Indels 0; Gaps 0;

692 CGTATCGACAGAAATGATCGCTGTGGTATGAGAGTTGTAAGTGCAGCGAGAAAA 751
:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
195 SDEBEVLPDIDVEDVDVDELNOEQVADLNKQATTYGMAADGDEVRLRKDKREAAELIKAKA 254
:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY      752 GTGGTCGACATGGCGAACCGTGACACCTGGGAAATCGCTGTGATGACGCGCGGGATC 811
Db      255 LEEBKAMYSGRRSRQRGRFRERKRLRKRTISPBYARDSPITYPYKRSBESSSESRSR 314

```

```
QY      812 CACAAGCGTTGGAAGTGTGCGGCAAAATTGAGTTGTTGATCGATGGAGCGTCGGCAATCG 871
      315 SRSPTPGRREKITTFTISFGGSDDEAAAAAASGVTTGKPPAPQPGPAPGNASAR 374
QY      872 CCGGCGCCCGGATGCGTGCCCGGAGCTGCTTGTGGGAGACGATCAACGATGCTCAGA 931
      375 RISSSSSSSSASRTSSSSSSSRSGRRGGYTRSGRHARSRSRWSSRSRRYSR 434
Db      375 RISSSSSSSSASRTSSSSSSSRSGRRGGYTRSGRHARSRSRWSSRSRRYSR 434
QY      932 TCGTTGACGAGAAAGATCTGCCCGCATGATGATTCATTCGCGCGCATCGTGA 991
      435 SRSRGRRHSGGSDGHRYSRSPARFGYGPRRRSRSHSGDYYRGGRLRHSSSRS 494
Db      435 SRSRGRRHSGGSDGHRYSRSPARFGYGPRRRSRSHSGDYYRGGRLRHSSSRS 494
QY      992 TGATGATTGGAAGCTTGTGCGCCTTACGTTTGGCGAGATCCTCGACTGTCGAAGCTG 1051
      495 RSSWSLSPSRSLTRSRSHSPSQSRSSRSRSPSPSPAREKLTTPAASPAVGEKL 554
Db      495 RSSWSLSPSRSLTRSRSHSPSQSRSSRSRSPSPSPAREKLTTPAASPAVGEKL 554
QY      1052 GCGGAGCTTGGGAAAATCTTCGTGA 1076
      555 KTEBPAGKETGAAXVTQADASGEA 579
Db      555 KTEBPAGKETGAAXVTQADASGEA 579
```

```
RESULT 15
US-10-123-155-10
; Sequence 10, Application US/10123155
; Publication No. US20030068794A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Geritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gunney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3330R1C30
; CURRENT APPLICATION NUMBER: US/10/123,155
; CURRENT FILING DATE: 2002-04-15
; Prior Application removed - See Palm or File Wrapper
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 10
; LENGTH: 594
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-123-155-10
```

Query Match 3.3%; Score 37.6; DB 14; Length 594;
Best Local Similarity 9.4%; Pred. No. 0.12;
Matches 36; Conservative 145; Mismatches 204; Indels 0; Gaps 0;

```
QY      692 CGTATCTGACGAAATGCACTCGCGCTGTGATGAGAGTTGAGTGGCGACGAGAAAA 751
      195 SDEDEVYIPDIVDEVDELNQEQVADLNKQATTYGMADGDFVMLRKDEAEAIKHAKA 254
QY      752 GTGGTGCACATGCGCAACCGTGCACCGCTGGGAAATCGCTGTGATGACAGTGGCGGGGATC 811
      255 LBEKXAMYSGRRRSRQRREFRERLRGRKISPPYARBDSPYTPYKRSPSSSSRSR 314
QY      812 CACAAGCGTTGGAAGTGTGCGGAGAAATTGAGTTGATGATGAGAGCTCGGCAATCG 871
      315 SRSPTPGRREKITTFTISFGGSDDEAAAAAASGVTTGKPPAPQPGPAPGNASAR 374
QY      872 CGTGTCCCGTGGATTCGTCGCCGAGCTGCTTGGGGGAGAACTAACAGATGCTCAGA 931
```

```
Db      375 RISSSSSSSSASRTSSSSSSSRSGRRGGYTRSGRHARSRSRWSSRSRRYSR 434
QY      932 TCGTTGACGAGAAAGATCTGCCCGCATGATGATTCATTCGCGCGCATCGTGA 991
      435 SRSRGRRHSGGSDGHRYSRSPARFGYGPRRRSRSHSGDYYRGGRLRHSSSRS 494
Db      435 SRSRGRRHSGGSDGHRYSRSPARFGYGPRRRSRSHSGDYYRGGRLRHSSSRS 494
QY      992 TGATGATTGGAAGCTTGTGCGCCTTACGTTTGGCGAGATCCTCGACTGTCGAAGCTG 1051
      495 RSSWSLSPSRSLTRSRSHSPSQSRSSRSRSPSPSPAREKLTTPAASPAVGEKL 554
Db      495 RSSWSLSPSRSLTRSRSHSPSQSRSSRSRSPSPSPAREKLTTPAASPAVGEKL 554
QY      1052 GCGGAGCTTGGGAAAATCTTCGTGA 1076
      555 KTEBPAGKETGAAXVTQADASGEA 579
Db      555 KTEBPAGKETGAAXVTQADASGEA 579
```

Search completed: November 8, 2003, 21:35:15
Job time : 313 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 8, 2003, 18:48:11 ; Search time 1964 Seconds
(without alignments)
14107.486 Million cell updates/sec

Title: US-10-007-527A-1

Perfect score: 1140

Sequence: 1 atgaccagcgtacgtgcctga.....taatacgggttcgcaagtga 1140

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST:*
1: em_estba:*
2: em_estlum:*
3: em_estlin:*
4: em_estlum:*
5: em_estlov:*
6: em_estpl:*
7: em_estro:*
8: em_hcc:*
9: gb_est1:*
10: gb_est2:*
11: gb_hcc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estlum:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_pro:*
26: em_gss_png:*
27: em_gss_vrt:*
28: gb_gss1:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	43	3.8	767	10	BF862931 963039F08
2	43	3.8	1071	29	CC304102 CH261-221
3	41.8	3.7	609	14	CD432432 ETH1.30 H
4	41.8	3.7	885	13	BK425603 BK425603

C 5	40.8	3.6	1201	13	BX394611	BX394611 BX394611
C 6	40.6	3.6	556	12	B1440943	B1440943 1c78e08.Y
C 7	40.6	3.6	737	12	BW933131	BW933131 UT-M-BH3-
C 8	40.6	3.6	744	13	BQ442682	BQ442682 UT-M-ETH-
C 9	40.6	3.6	937	13	BUS91391	BUS91391 AGENCOURT
C 10	40.6	3.6	1122	11	BK084367	BK084367 Mus muscu
C 11	40.2	3.5	1201	13	BX376097	BX376097 BX376097
C 12	39.4	3.5	1201	13	BX381961	BX381961 BX381961
C 13	38.8	3.4	759	9	A1663718	A1663718 ud47b04.Y
C 14	38.8	3.4	365	10	BG406185	CA361593 635452 NC
C 15	38.2	3.4	491	12	B1469223	BG406185 sac36104.
C 16	38.2	3.4	515	10	BQ404909	B1469223 sac109a10.
C 17	38.2	3.4	568	9	AW185913	BQ404909 sac42908.
C 18	38.2	3.4	568	9	AW185913	AW185913 sac42908.
C 19	38	3.3	773	12	B1555257	B1555257 sac1b10.Y
C 20	38	3.3	787	29	CNS016H3	B1555257 603236008
C 21	38	3.3	812	9	AL695087	AL106737 Drosophi1
C 22	37.8	3.3	686	28	BH576806	AL695087 AL695087
C 23	37.6	3.3	825	29	CNS02P5W	BH576806 BOCRK62TF
C 24	37.6	3.3	1402	10	BG395673	AL194621 Tetradon
C 25	37.4	3.3	639	10	BH658123	BG395673 602458312
C 26	37.2	3.3	1201	13	BX381961	BH658123 BH658123
C 27	36.8	3.2	636	12	BG820861	BX381961 BX381961
C 28	36.6	3.2	422	9	A1437648	BG820861 602780932
C 29	36.6	3.2	485	10	BG650579	A1437648 sac37g10.Y
C 30	36.6	3.2	595	14	CA785996	BG650579 sac36g01.
C 31	36.6	3.2	639	10	BG415045	CA785996 sac42a08.
C 32	36.6	3.2	1101	29	CNS016TW	BG415045 HVSMEK00
C 33	36.4	3.2	692	12	BMS96227	AL106810 Drosophi1
C 34	36.2	3.2	374	12	B1780520	BMS96227 170006874
C 35	36.2	3.2	528	14	CA029424	B1780520 EBes01.SQ
C 36	36.2	3.2	628	13	BQ739865	CA029424 HZ65C19F
C 37	36.2	3.2	334	9	AW508862	BQ739865 HB04C07.H
C 38	36	3.2	925	29	CNS0091P	AW508862 s141c07.Y
C 39	36	3.2	1201	13	BX444391	AT050313 Drosophi1
C 40	35.8	3.1	292	10	BA431085	BX444391 BX444391
C 41	35.8	3.1	645	29	CNS01213	BA431085 BA431085
C 42	35.6	3.1	1163	9	AL577643	AL101589 Drosophi1
C 43	35.6	3.1	317	14	CB243032	AL577643 AL577643
C 44	35.6	3.1	637	10	BB581130	CB243032 UT-CF-FNO
C 45	35.6	3.1	849	13	BX462111	BB581130 BB581130
						BX462111 BX462111

ALIGNMENTS

RESULT 1
LOCUS BF862931
DEFINITION 963039F08.Y1 C. reinhardtii CC-1690, Stress condition I, normalized
ACCESION BF862931
VERSION BF862931.1 GI:12253075
KEYWORDS
SOURCE
ORGANISM Chlamydomonas reinhardtii
Chlamydomonas reinhardtii
Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
Chlamydomonadaceae; Chlamydomonas.

REFERENCE
AUTHORS Grossman, A., Davies, J., Federspiel, N., Harris, E., Hauser, C.,
Lefebvre, P., McDermott, J. P., Shigeru, J., Sillflow, C., and Stern, D.
TITLE Analyses of the Chlamydomonas reinhardtii Genome: A Model,
Unicellular System for Analyzing Gene Function and Regulation in
Vascular Plants: project phase 3
JOURNAL Unpublished
COMMENT Contact: Charles Hauser
Duke University
Durham, NC 27708-1000
Tel: 919 613 8159
Fax: 919 613 8177
Email: chauser@duke.edu.
Location/Qualifiers

FEATURES

```

source
1. .767
  /organism="Chlamydomonas reinhardtii"
  /mol_type="mRNA"
  /strain="CC-1690 wild type mt+ 21gr"
  /db_xref="taxon:3055"
  /clone_lib="C. reinhardtii CC-1690, stress condition I,
normalized, lambda zap II"
  /note="vector: pluescript II SK-, Site 1: EcoRI, Site 2:
XhoI; This library, constructed by John Davies and Jeffrey
McDermott, combines cDNAs from CC-1690 cells grown to
mid-log phase in TAP-N (30 min, 1hr, 4hr), TAP-S (30 min,
1hr, 4hr), TAP-P (4hr, 12hr, 24hr), NO3 to NH4 (30min, 1hr,
4hr) and NH4 to NO3 (30min, 1hr, 4hr). PolyA mRNA was
purified from each sample, pooled and cDNA synthesized.
The cDNA was directionally cloned into lambda zap II
(Stratagene) in the EcoRI (5') and XhoI (3') sites.
pluescript II SK- plasmids were excised from the lambda
zap clones by superinfection with Exsist (Stratagene)
phase. The library was normalized using method 4 described
in Donaldo et al (1996) Genome Research 6: 791-806."
133 a 238 c 288 g 107 t 1 others
BASE COUNT
ORIGIN

```

Query Match	3.8%	Score 43	DB 10	Length 767
Best Local Similarity	48.6%	Pred. 0.26		
Matches 118	Conservative 0	Mismatches 125	Indels 0	Gaps 0
QY	390	GGCAGCCTGGAAAGCTGGACCAACGTCGTCGTGGCGCTACCGAAGCGTAAATGTACGG	449	
Db	190	GGCGGCTTCGAGAGCGGAGGCTGGCCGCGATTATGTCGGCGCTGGAGACGCCAAGAACCG	249	
QY	450	CTGGAGACGGATACGTCGCGCGCTGTTGAAATACCTCAGCGAAAAACGCTGGCAGCTTCA	509	
Db	250	CGGGAGGACTACTCGCTCTGGCAGATATCGCGCGCTGGAGAAAGCGTTCAGAGAGGCGCA	309	
QY	510	CGTTCACGCGCTACTCATGTCATGTCACAGTGTGACGTGATAGAAACAATCTCCAATCCTTCTC	569	
Db	310	CGTAAAGCTGAAAGCCCAAGACGATGGGGCGCCGCAACCCATCTTCCGGGCGCGCTGTGGA	369	
QY	570	GGATGCGATGTTTCATTCGCTGGAATCTCCAAACTGATCTCTGGGATTTGCTGGCGCACT	629	
Db	370	GGCGGCGGTGGCGGCGGCGCGCAGGCTCGCACTTCGAGTCGCTGATGAGGCTTCGCGCC	429	
QY	630	ACG 632		
Db	430	GCG 432		
RESULT 2	CC304102	1071 bp	DNA	linear
LOCUS	CC304102	CH261-22111	Sp6.1	CH261
DEFINITION	CH261-22111	Sp6.1	Gallus gallus	genomic clone CH261-22111,
ACCESSION	CC304102			genomic survey sequence.
VERSION	CC304102.1	GI:30675543		
KEYWORDS	GSS.			
SOURCE	Gallus gallus	(chicken)		
ORGANISM	Gallus gallus			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
AUTHORS	Archaeosauria; Aves; Neognathae; Galliformes; Phasianidae;			
TITLE	Phasianidae; Gallus.			
JOURNAL	1 (bases 1 to 1071)			
COMMENT	Kremlicki, C., Higginbotham, J., Wylie, K., Carter, J., McPherson, J.,			
	Warren, W., Graves, T., Mardis, E., and Wilson, R.			
	Gallus gallus BAC End Reads			
	Unpublished			
	Contact: Richard K. Wilson			
	Genome Sequencing Center			
	Washington University School of Medicine			
	Email: submissions@wustl.edu			
	Insert Length: 182000	Std Error: 0.00		
	Seq primer: Sp6 ATTTAGGTGACACTATAG			
	Class: BAC ends			

FEATURES	source
High quality sequence start: 27	
High quality sequence stop: 476.	
Location/Qualifiers	
1..1071	
/organism="Gallus gallus"	
/mol_type="genomic DNA"	
/strain="Red Jungle Fowl"	
/db_xref="taxon:9031"	
/clone="CH261-22111"	
/sex="female"	
/cell_line="UCD001, inbred 256"	
/clone_1b="CH261"	
/note="Vector: PTARBAC2.1; Site.1: EcoRI; Site.2: EcoRI; CH61 Female Chicken library - For library and clone ordering information: http://www.chori.org/bacpac "	
BASE COUNT	254 a 312 c 213 g 290 t 2 others
ORIGIN	
Query Match	3.8%; Score 43; DB 29; Length 1071;
Best local Similarity	53.2%; Pred. No. 0.31;
Matches	91; Conservative 0; Mismatches 80; Indels 0; Gaps 0;
Qy	840 TGAAGTTGGTTCATGGAGAGCTGGGCGCAATCGCTGGTCCCGTGAATTCGCTGCCGAGC 899
Db	
Qy	69 TGTGTGTGCTGCGTGGGAGCTGTGCAGAGAGGTGAATTCACCCCGATGTGTGCAAGTGC 128
Db	
Qy	900 TGTGCTTGGGGCAGAACTAACAGATGCTCAATGTTGAGCAGAGGAAGATCTGCCCGGT 959
Db	
Qy	129 TGGCACTGCATTAACATGCTGCTAATGTGCTCGTGGCTGGGCAAGAAAGCTTGAATGCA 188
Db	
Qy	960 CATGGTTGCATCATTCGCGGGCGAATGCTGGATGATGATTCGCACTTGTGC 1010
Db	
Qy	189 CCGGGTCTGTTACAGCTGTATGACTCTTGAGACCGGTAGCTGAGACTGTGC 239
Db	

RESULT 3				
CD432432				
LOCUS	CD432432	609 bp	mRNA	linear EST 03-JUN-2003
DEFINITION	ETH1_30_H03.b1.A002 Ethylane-treated seedlings Sorghum bicolor cDNA			
ACCESSION	CD432432			
VERSION	CD432432.1	GI:31348075		
KEYWORDS	EST.			
SOURCE	Sorghum bicolor (sorghum)			
ORGANISM	Sorghum bicolor			
REFERENCE	Eukaryote; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoideae; Andropogoneae; Sorghum.			
AUTHORS	1 (bases 1 to 609) Cordonnier-Pratt,M.-M., Wentzel,V., Suzuki,Y., Sugano,S., Klein,R.R., Liang,C., Sun,F., Sullivan,R., Shah,M., Salzman,R., Chua Tair,N., Gonzalez,M., Lane,S., Miller,V., Nanda,F., Olaseinde,O., Eastman,A. and Pratt,L.H. An EST database from Sorghum: 1-aminocyclopropane-1-carboxylic acid (ACC)-treated seedlings			
TITLE	Unpublished			
JOURNAL	Other ESTs: ETH1_30_H03.g1.A002			
COMMENT	Contact: Cordonnier-Pratt MM Laboratory for Genomics and Bioinformatics The University of Georgia, Department of Plant Biology Plant Sciences Building, Km. 2502, Athens, GA 30602-7271, USA Tel: 706 542 1860 Fax: 706 583 0210 Email: mmpratt@uga.edu Library constructed by Dr. Yutaka Suzuki and Dr. Sumio Sugano in the Human Genome Center, University of Tokyo Institute of Medical Science; plant material and RNA prepared at Texas A & M University, sequencing done in the laboratory for Genomics and Bioinformatics, University of Georgia. Sequence ends have been trimmed to exclude vector and regions below phred quality 16. Three-prime sequences are presented as their reverse complement and have been trimmed to exclude polyA. Seq primer: Sng3 (CGACCTGCAGCTCGACGACA)			

TITLE Full-length cDNA libraries and normalization

UI-M-BH3-bw-a-11-0-UI 5', mRNA sequence.

ACCESSION
BM333131 GI:19392283
VERSION
BM333131.1
KEYWORDS
EST
SOURCE
Mus musculus (house mouse)
ORGANISM
Mus musculus

REFERENCE
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 737)
AUTHORS
Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE
Normalization and subtraction: two approaches to facilitate gene
discovery

JOURNAL
MEDLINE
PUBMED
97044477
8889548
COMMENT
Contact: Chin, H
National Institute of Mental Health
6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD
20892-9643, USA
Tel: 301 443 1706
Fax: 301 443 9890
Email: m85t@mail.nih.gov
CDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
CDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Researchers may obtain clones from Research
Genetics (www.resgen.com).
Seq primer: M13 REVERSE.

FEATURES

Location/Qualifiers
1..737
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UI-M-BH3-bw-a-11-0-UI"
/dev_stage="27-32 days"
/lab_host="DH10B (Life Technologies)"
/clone_id="NIH_BMAP_M_S4"
/note="Vector: pYT73D-Pac (Pharmacia) with a modified
polylinker; Site 1: Not I; Site 2: Eco RI; The
NIH_BMAP_M_S4 library is a subtracted library of a series,
ultimately derived from a mixture of individually tagged
normalized libraries from ten regions of the mouse brain
(cerebellum, brain stem, olfactory bulbs, hypothalamus,
cortex, amygdala, basal ganglia, pineal gland, striatum,
hippocampus) after a series of subtractions to reduce the
representation of cDNAs from which ESTs had already been
generated. The following serially subtracted libraries
were generated in this process: NIH_BMAP_M_S4,
NIH_BMAP_M_S3.3, NIH_BMAP_M_S3.2, NIH_BMAP_M_S3.1,
NIH_BMAP_M_S2, NIH_BMAP_M_S1. The subtracted library
(NIH_BMAP_M_S4) was constructed as follows: PCR amplified
cDNA inserts from NIH_BMAP_M_S3.3, NIH_BMAP_M_S3.2, and
NIH_BMAP_M_S3.1 clones from which 3' ESTs had been derived
was used as a driver in a hybridization with a pool of
the NIH_BMAP_M_S3.3, NIH_BMAP_M_S3.2, and NIH_BMAP_M_S3.1
libraries in the form of single-stranded circles. The
remaining single-stranded circles (subtracted library)
was purified by hydroxyapatite column chromatography
converted to double-stranded circles and electroporated
into DH10B bacteria (Life Technologies) to generate the
NIH_BMAP_M_S4 library. This procedure has been previously
described (Bonaldo, Lennon and Soares, Genome Research
6:791-806, 1996)"

BASE COUNT
ORIGIN
122 a 246 c 218 g 149 t 2 others

Query Match 3.6%; Score 40.6; DB 12; Length 737;
Best Local Similarity 48.5%; Pred. No. 1.3;
Matches 112; Conservative 0; Mismatches 119; Indels 0; Gaps 0;

QY 696 TCTGACGAATTCGATCTGCGCTTGTATGAGCTGTGATGCGCACGAAAAAGTGG 755

Db 564 TCCTGAGAGAGCTGACGCGCTGCGGCGAGAGCTGGGCGGCTGGCGGCGATGCGGA 505
QY 756 TCACATGCGCAACCGTCGACCCCTGGAATCGCTGTGATGATGATGAGCGGGATCCACA 815
Db 504 GCGCGTGCAGACCGCGTGGCGCCCGCAGACCGCAGCTGTGTGGCGCGCTGGAGCCGCTGC 445
QY 816 AGCGTTGGAATGTTGCGGAGATTGATGTTGTTGATGAGGACGTCGCGCATCGCGTG 875
Db 444 TGCAGAGAGACCGTGAAGCGAGCTTACGAGCGCTGGCGCCCTGGAAGAGCGGAGCGCGGC 385
QY 876 GTCCCGTGAATGCTGCTGCGGAGCTGCTTGGGCGAGAACTAACAGATGC 926
Db 384 GACCCGAGGCGCAGTGCCTGAGCGCTGTGAGAGAACTCGCGC 334

RESULT 8
LOCUS
DEFINITION
BQ442682 744 bp mRNA linear EST 29-MAY-2002
UI-M-EV0-bxf-a-13-0-UI.r1 NIH_BMAP_EV0 Mus musculus CDNA clone
IMAGE:5707284 5', mRNA sequence.
ACCESSION
BQ442682
KEYWORDS
BQ442682.1 GI:21245794
SOURCE
EST
Mus musculus (house mouse)
ORGANISM
Mus musculus

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
Contact: Robert Strauberg, Ph.D.
Email: cgaabs-remail.nih.gov
Tissue Procurement: Dr. James Lin, University of Iowa
CDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
CDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.jnl.gov
This clone was contributed by the Brain Molecular Anatomy Project
(BMAP)

FEATURES

Location/Qualifiers
1..744
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6"
/db_xref="taxon:10090"
/clone="IMAGE:5707284"
/tissue_type="whole brain"
/dev_stage="embryo 15.5 dpc"
/lab_host="DH10B (Life Technologies)"
/clone_id="NIH_BMAP_EV0"
/note="Organ: Brain; Vector: pYX-Ase; Site 1: Eco RI;
Site 2: Not I; The library was constructed according to
Bonaldo, Lennon and Soares, Genome Research, 6:791-806,
1996. Denatured mRNA was size fractionated on a 1% agarose
gel. First strand cDNA synthesis was primed with an
oligo-dT primer containing a Not I site. Double stranded
cDNA was size selected according to mRNA size fraction,
ligated with EcoR I adaptor, digested with Not I, and then
cloned directionally into pYX-Ase vector. The library tag
sequence located between the Not I site and the polyA tail
, is GTGCGTGGAA. This library was created for the
University of Iowa Mouse Brain Molecular Anatomy Project
(BMAP). 'Gene Discovery in the Developing Mouse Nervous
System', supported by National Institutes of Mental Health
(NIMH), Hemlin Chin, Ph.D., program coordinator."

BASE COUNT
ORIGIN
140 a 212 c 244 g 146 t 2 others

Query Match 3.6%; Score 40.6; DB 13; Length 744;

Db 562 TCCGTGAGAGCTGCACGGCTGCAGGCGAGACTGGGGCGGCTGGCGGGCGGCATGGCGA 503

756 TCGACATGGCAACCGTCACCCCTGGGAAATCCGTGTGATGCAGTGGGCGGGGATCCACA 815

DB 502 GGCCTGCGCAGCCGGTGGCCCGCCAGACGCGCTGGAGCCCGTGC 443

[illegible]

876 GTCCCGTGAA TTGCGTCCGAGCGTAGTCTTTGGGGCAGAACTTAA CAGATGC 925

Db 382 GACCCGAGGCCACAGTGCCTGGCCTTAGGGCCGTGTGCTGGAGAGAACTGCGGC 332

3
1
3
4
3
5
5

AK084367	1122 bp	mpv	1:1000	TIME OF DNA CON
AK084367/c				
LOCUS				
AK084367				

DEFINITION Mus musculus 12 days embryo eyeball cDNA, RIKEN full-length enriched library, clone:D230032H18 product: pentaxin related gene.

full insert sequence.
ACCESSION AK084367

AK08436/1.1 GI:26101923
HTC; CAP trapper.

ORGANISM: Mus musculus (mouse)
Phylum: Metazoa: Chordata: Vertebrata: Euteleostomi:

REFERENCE
1 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

AUTHORS	Carninci, P. and Hayashizaki, Y.
TITLE	High-efficiency full-length cDNA cloning

00000000
 00000000
 MEDLINE
 99279253
 10240536
 00000000

REFERENCE
2
Carninci P., Shibata Y., Havatsu N., Suwahara Y., Shibata K.

TITLE Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y. Normalization and subtraction of cap-trapper-selected cDNAs to

prepare full-length cDNA libraries for rapid discovery of new genes
Genome Res. 10 (10), 1617-1630 (2000)

INDEXED	202557
PUBMED	11042159
REFERENCE	3

AUTHORS
Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,
Konno, H., Akiyama, J., Nishi, K., Kitsumi, T., Tashiro, H., Itoh M.

Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.

Fujiwaka, S., Inoue, K., Iogawa, Y., Izawa, M., Ohara, E., Wataniki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura, S., Kawai, J., Ohno, S.

TITLE	INSTRUMENT	ANALYST	DATE	TIME	FILE NAME
RIKEN integrated sequence analysis (RISA) system--384-format semencing pipeline with 384 multiplexed primer pairs	ABI PRISM 3700XL	J. L. RIKEN	10/10/2006	10:00	RISA_384_format_10-10-2006_10-00-00

JOURNAL Genome Res. 10 (11), 1757-1771 (2000)
MEDLINE 20530913

POBMED	11076861
REFERENCE	4

Amaki, Y., Onizugawa, H., Shimada, N., Ioshino, M., Itoh, M., Ishii, T.,
Aizawa, T., Hara, A., Fukunishi, Y., Komno, H., Adachi, J., Fukuda, S.,
Aizawa, K., Tera, M., Nishi, Y., Watanabe, H., Kashi, C., Watanabe, T.

Saito, T., Okazaki, Y., Gojibori, T., Bono, H., Kasukawa, T., Saito, R., Kadota, K., Matsuda, H., Ashburner, M., Batilov, S., Casavant, T.

Fleischmann, W., Gaasterland, T., Gissi, C., King, B., Kochiwa, H., Kuehl, P., Lewis, S., Matsuo, Y., Nakaido, I., Pesole, G.

Quackenhush, U., Schriml, L.M., Staabli, F., Suzuki, R., Tomita, M.,
Wagner, L., Washio, T., Sakai, K., Okido, T., Furuno, M., Aono, H.,

Arcuate, K., Basmis, S., Brave, V., Bollelli, V., Bojunga, N.,
Carninci, P., de Bonaldo, M.F., Brownstein, M.J., Bulc, C.,
Fletcher, C., Fujita, M., Gariboldi, M., Grottelich, C., Hui, D.

Hofmann, M., Humé, D. A., Kamiya, M., Lee, N. H., Lyons, P.,
Marchionni, L., Mashima, J., Mazzarelli, J., Mombaerts, P., Nordone, P.

Ring B., Ringwald, M., Rodriguez, I., Sakamoto, N., Sasaki, H., Sato, K., Schonbach, C., Seya, T., Shibata, Y., Storch, K.F., Suzuki, H.

JOY-O-UNA, A., MANG, N.H., WEILZ, C., WHITEHEAD, C., WILLING, L.,

Ynshaw-Boris, A., Yoshida, K., Hasegawa, Y., Kawaji, H., Kohsaki, S., and Hayashizaki, Y.
Functional annotation of a full-length mouse cDNA collection
JOURNAL
Nature 409 (6821), 685-690 (2001)
MEDLINE
21085660
PUBMED
11217851

5
The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.
Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
JOURNAL
Nature 420, 563-573 (2002)
6 (bases 1 to 1122)

ADACHI, J., AIZAWA, K., AKIMURA, T., ARAKAWA, T., BONO, H., CARINCI, P., FUKUDA, S., FURUKO, M., HANAGAKI, T., HARA, A., HASHIZUME, W., HAYASHIDA, K., HAYATSU, N., HIRAMOTO, K., HIRAOKA, T., HIROZANE, T., HORI, F., IMOTANI, K., ISHII, Y., ITOH, M., KAGAWA, I., KASUKAWA, T., KATO, H., KAWAI, D., KOJIMA, Y., KONDO, S., KONNO, H., KODA, M., KOYA, S., KURIHARA, C., MATSUYAMA, T., MIYAZAKI, A., MURATA, M., NAKAMURA, M., NISHI, K., NOMURA, K., NUMAZAKI, R., OHNO, M., OHSAO, N., OKAZAKI, Y., SAITO, K., SAITOH, H., SAKAI, C., SAKAI, K., SAKAZUME, N., SANO, H., SASAKI, D., SHIBATA, K., SHINAGAWA, A., SHIRAKI, T., SOGABE, Y., TAGAMI, M., TAGAWA, A., TAKAHASHI, F., TAKAKU-AKHIRA, S., TAKEEDA, Y., TANAKA, T., TOMARU, A., TOYA, T., YASUNISHI, A., MURAMATSU, M., and HAYASHIZAKI, Y.
Direct Submission
Submitted (16-Apr-2002) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-resgsc.riken.go.jp, URL: http://genome.resgsc.riken.go.jp, Tel: 81-45-503-9222, Fax: 81-45-503-9216)

COMMENT
cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.
Please visit our web site for further details.
URL: http://genome.resgsc.riken.go.jp/
URL: http://fantom.resgsc.riken.go.jp/
Location/Qualifiers
1..1122
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="PANTOM:DB:D230032H18"
/db_xref="taxon:10090"
/clone="D230032H18"
/tissue_type="eyeball"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="12 days embryo"
1..1122
/note="Pentaxin related gene (MGI:104641, GB|NM_008987, evidence: BLASTN, 100%, match=404)"

BASE COUNT
244 a 327 c 292 g 259 t

ORIGIN
misc_feature

Query Match 3.6%; Score 40.6; DB 11; Length 1122;
Best Local Similarity 48.5%; Pred. No. 1.6;
Matches 112; Conservative 0; Mismatches 119; Indels 0; Gaps 0;

QY 696 TCTGACGAAATTGCTGCGCTGGTGTGAGGTTGAGTGGCGGACGGAAGTGG 755
DB 647 TCCGTGAGAGCTCAGCGCTGCGGCGAGAGCTGCGGCGCGGCGGCGATGGGA 588

QY 756 TCGACATGCAACGCTGACCGCTGGAATCGCTGTTGATGACATGGCGGGATCCACA 815
DB 587 GCGCGTGGCGAGCGCGTGGCGCCGCAAGCGCTGTGTGGCGGCGCTGAGCGCGTGC 528

QY 816 AGCGTTGAACTGTGGGAGAAATTGATTGTTGCTGATGGGACGTCGGGCAATCGGCTG 875
DB 527 TCGAGAGAGCGGTGAGCGAGCGCTCAGGCGTGGCGCGCTGGAAGACGCGAGCGCGGC 468

QY 876 GTCCGTGATTCGTCGCCGAGCTGCTGTTGGGCGAGAACTPACAGATGC 926
DB 467 GACCCGAGCGCAGAGTGCCTGAGCGCTGTGCTGAGAGAACTGGCGC 417

RESULT 11
BX376097 1201 bp mRNA linear EST 08-MAY-2003
LOCUS BX376097
DEFINITION BX376097 Homo sapiens NEUROBLASTOMA COT 25-NORMALIZED Homo sapiens cDNA clone CS0DC022M12 5-PRIME, mRNA sequence.
ACCESSION BX376097
VERSION BX376097.1 GI:30434756
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrate; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 1201)
Li, W.B., Gruber, C., Jeessee, J., and Polayes, D.
Full-length cDNA libraries and normalization
Unpublished
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: segret@genoscope.cns.fr, Web: www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of Invitrogen. This sequence belongs to sequence cluster 2866.f
Contact: Feng Liang Email: fliang@lifestech.com URL: http://fulllength.invitrogen.com/Invitrogen Corporation 1600 Faraday Avenue genoscope sequence ID: CS0DC022B060P1.
Location/Qualifiers
1..1201
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DC022M12"
/tissue_type="NEUROBLASTOMA COT 25-NORMALIZED"
/clone_lib="Homo sapiens NEUROBLASTOMA COT 25-NORMALIZED"
/note="1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with NotI and cloned into the NotI and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized."

BASE COUNT
79 a 152 c 66 g 270 t 634 others

ORIGIN

Query Match 3.5%; Score 40.2; DB 13; Length 1201;
Best Local Similarity 8.8%; Pred. No. 2.2;
Matches 54; Conservative 242; Mismatches 317; Indels 2; Gaps 1;

QY 165 TGTGACCATTTGCAACGCTCGAAGAGTTGATTCGAGAGCGCTGCTCGCGGAA 224
DB 431 TMTNNMCMCTTNTTNTTKTTTNTTKKNTTKKCKKMKKMKMKKNNCKKTKTKK 490

QY 225 GGGCTGATCTGCCCTGCTGTGGGAGAAAGTGCGACATCGTGCAGGAATTTC 284
DB 491 NKKNNKNNKNNNNNNKNNKNNKNNKNNKNNKNNKNNKNNKNNKNNKNNKNN 550

QY 285 TCAAGTTGTTGTCATCAACTCGGAGATGATCTGTTGGATGTGTACAGATGACCG 344
DB 551 KKKKKNN 610

QY 345 CCAATACAGCTGTCAGCGGCTCCAGACCTATGAGCTTTCGGAAGCTGGAAAGC 404
DB 611 CKKKNN 670

QY 405 TCGACCAAGCGTGTGCTGTTGCGTACGGAAGTGA--ATGTACGCGCTGCGAGTAC 462
DB 671 MKNN 730

QY 463 GTGGCGCGCTTTGAATCACTCGAGAAAGAGCGCTGCGACGTTCCAGCGCTGA 522
DB 731 KKKKKNN 790

Query Match 3.4%; Score 38.8; DB 9; Length 759;
 Best Local Similarity 52.1%; Pred. No. 4.4;
 Matches 85; Conservative 0; Mismatches 78; Indels 0; Gaps 0;

QY 764 GCACCGTGCACCTCGGAAATGCGTGTGATGAGTGGCGGAGATCCAGAGCTTGG 823
 DB 602 GCAGCGGTGGCCNCGAGACGAGTGTGTGGGGCGCTGGAGCCGCTGCTGCAGAG 543
 QY 824 AACTGTGGCGAATTGATTGATTGATGAGCGTGGGCGAATGCGGTGTCGCG 883
 DB 542 AGCCCTGACGAGGCTTACGAGCTGCGCGCTGAGAGACGAGGCGCGCAACCGAG 483
 QY 884 GATTGCGGCGGAGCTGCTTGGGCGAAGTACAGATGC 926
 DB 482 GCGACAGTGCCTGGCTGAGCGCTGTGTGTGAGGAGATGCGGC 440

RESULT 14
 CA361593 777 bp mRNA linear EST 06-NOV-2002
 LOCUS 63452 NCCWA 1RT Oncorhynchus mykiss cDNA clone 1RT6015_A_H08 5';
 DEFINITION mRNA sequence.

ACCESSION CA361593.1 GI:24667466
 VERSION CA361593
 KEYWORDS EST.
 ORGANISM Oncorhynchus mykiss (rainbow trout)
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Actinopterygii; Neopterygii; Teleostei; Euteleostei;
 Proclanchiiformes; Salmoniformes; Salmonidae; Oncorhynchus.

REFERENCE 1 (bases 1 to 777)
 AUTHORS Rexroad, C.E. and Keele, J.W.
 TITLE Sequence analysis of a rainbow trout normalized cDNA library
 JOURNAL Unpublished
 COMMENT Contact: Rexroad CE
 USA, ARS, National Center for Cool and Cold Water Aquaculture
 11876 Leetown Road, Kearneysville, WV 25430, USA
 Tel: 304 724 8340 X2129
 Fax: 304 725 0351
 Email: crexroad@nccwa.ars.usda.gov

FEATURES
 source
 1..777
 Location/Qualifiers
 /organism="Oncorhynchus mykiss"
 /mol_type="mRNA"
 /db_xref="taxon:8022"
 /clone="1RT6015_A_H08"
 /tissue_type="pooled"
 /lab_host="DH10B"
 /clone_id="NCCWA 1RT"
 /note="Vector: pCMV SPORT6; Site 1: NotI; Site 2: SalI;
 library made from pooled tissue from brain, gill, liver,
 spleen, muscle, and kidney."

BASE COUNT 199 a 221 c 188 g 169 t
 ORIGIN

Query Match 3.4%; Score 38.8; DB 14; Length 777;
 Best Local Similarity 54.1%; Pred. No. 4.4;
 Matches 79; Conservative 0; Mismatches 67; Indels 0; Gaps 0;

QY 282 TTCTCAAGTGTGCTCATCACTGGGACTGATCTTGGCATGTGTGACGATGACAT 341
 DB 285 TTCTCAAGTGTGCTCATCACTGGGACTGATCTTGGCATGTGTGACGATGACAT 344
 QY 342 GCGCCATACAGCTGTGTCAGCGCTTCAGACCTTATGACTGACTTTCGACGCTGAA 401
 DB 345 TCCCTTCCGGCTCCGAGGGTCCCAAGGCTTAGGGGTGTATGACAGAGATTTCT 404
 QY 402 AGCTGCGACCAAGCTGCTGTTGC 427

DB 405 CGCTGCTCCCGCAGCAAGTCCGTGC 430

RESULT 15
 BG406185/c 365 bp mRNA linear EST 28-NOV-2001
 LOCUS sac36f04.y1 Gm-cl051 glycine max cDNA clone GENOME SYSTEMS CLONE
 DEFINITION ID: Gm-cl051-4472 5' similar to TR:004253 004253 T10M13.14.; mRNA
 sequence.

ACCESSION BG406185.1 GI:33312534
 VERSION BG406185.1
 KEYWORDS EST.
 ORGANISM Glycine max (soybean)
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
 ; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseolaeae;
 Glycine.

REFERENCE 1 (bases 1 to 365)
 AUTHORS Shoemaker, R., Keim, P., Vodkin, L., Erpelting, J., Coryell, V., Khanna,
 A., Bolla, B., Maira, M., Hillier, L., Kucaba, T., Martin, J., Beck, C.,
 Wylie, T., Underwood, K., Stepcie, M., Theising, B., Allen, M., Bowers,
 Y., Person, B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk,
 R., Ratter, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann,
 R., Waterston, R. and Wilson, R.

TITLE Public Soybean EST Project
 JOURNAL Unpublished
 COMMENT Contact: Shoemaker R/Public Soybean EST Project
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@wustl.edu
 This clone is available through: Reggen, Invitrogen Corp. 2130
 South Memorial Parkway Huntville, AL 35801 for further information
 call: (800)-533-4363 or contact via email: cculeregen.com
 High quality sequence scop: 342.

FEATURES
 source
 1..365
 Location/Qualifiers
 /organism="Glycine max"
 /mol_type="mRNA"
 /db_xref="taxon:3847"
 /clone="GENOME SYSTEMS CLONE ID: Gm-cl051-4472"
 /tissue_type="floral meristematic mRNA"
 /lab_host="DH10B"
 /clone_id="Gm-cl051"
 /note="Vector: pBluescript II SK+; Site 1: EcoRI; Site 2:
 XhoI; The cDNA library was constructed from floral
 meristematic mRNA provided by Dr. Halina Knap of Clemson
 University. Complementary DNA was synthesized from mRNA
 using a primer consisting of a poly(dT) sequence with a
 XhoI restriction site. EcoRI adapters were ligated to the
 blunt-ended cDNA fragments followed by XhoI digestion. The
 cDNA fragments were directionally cloned into the
 EcoRI-XhoI restriction site of the pBluescript vector. The
 ligated cDNA fragments were transformed into DH10B host
 cells (GibcoBRL). This library was constructed in the
 laboratory of Dr. Randy Shoemaker."

BASE COUNT 75 a 154 c 57 g 78 t
 ORIGIN

Query Match 3.4%; Score 38.2; DB 10; Length 365;
 Best Local Similarity 53.7%; Pred. No. 4.4;
 Matches 79; Conservative 0; Mismatches 68; Indels 0; Gaps 0;

QY 681 AGTTTCGCTGCGATCTGACGAAATTCATCTGGCGTGTATGAGGTTGTAAGTG 740
 DB 261 AGACCTGAAGTTCAATTGAGGGAAGTGATTTGACAGAGGATTCAGAGTTGTTGG 202
 QY 741 CGACGAAATAGTGTGACATGGAACCGGACCCCTGGGAATTCGTTGATGACGT 800
 DB 201 AGACGAGGAAGTGAAAGATGTTCTCCGGGACCTGGGAATGTTGAGATGAGTGA 142

Qy 801 GGGGGGATCCACAAGCTTGGACT 827
||| |
Db 141 CGGCGCGATGAGCCCGGAGGACT 115

Search completed: November 8, 2003, 21:28:16
Job time : 1969 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 22, 2003, 09:04:13 ; Search time 1957 Seconds

(without alignments)
14157.947 Million cell updates/sec

Title: US-10-007-527A-1
Perfect score: 1140
Sequence: 1 atgacagcagcgaagtcgtga.....taatacgcgtccgaagtrga 1140

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 22781392 seqs, 1215238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
EST:*
1: em_estba:*
2: em_esthum:*
3: em_estlin:*
4: em_estmu:*
5: em_estov:*
6: em_estro:*
7: em_estro:*
8: em_hnc:*
9: gb_est1:*
10: gb_est2:*
11: gb_est3:*
12: gb_est4:*
13: gb_est5:*
14: gb_est6:*
15: em_esthum:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vit:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rtd:*
26: em_gss_phg:*
27: em_gss_vil:*
28: gb_gss1:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	43	3.8	767	10	BF662931 963039P08
2	43	3.8	1071	29	CC304102 CH261-221
3	41.8	3.7	609	14	CD432432 ETH1_30.H
4	41.8	3.7	885	13	BX425603 BX425603

C 5	40.8	3.6	1201	13	BX394611 BX394611
C 6	40.6	3.6	556	12	B1440943 B1440943
C 7	40.6	3.6	737	12	BM933131 BM933131
C 8	40.6	3.6	744	13	BQ442682 BQ442682
C 9	40.6	3.6	937	13	B0591391 B0591391
C 10	40.6	3.6	1122	11	AK084367 AK084367
C 11	40.2	3.5	1201	13	BX376097 BX376097
C 12	39.4	3.5	1201	13	BX381961 BX381961
C 13	38.8	3.4	759	9	AL663718 AL663718
C 14	38.8	3.4	777	14	CA351593 CA351593
C 15	38.2	3.4	365	10	BG406185 BG406185
C 16	38.2	3.4	491	12	B1469223 B1469223
C 17	38.2	3.4	515	10	BG404909 BG404909
C 18	38.2	3.4	568	9	AW185913 AW185913
C 19	38	3.3	773	12	B155257 B155257
C 20	38	3.3	787	29	CNS016H3 CNS016H3
C 21	38	3.3	812	9	AL695087 AL695087
C 22	37.8	3.3	686	28	BH576806 BH576806
C 23	37.6	3.3	825	29	CNS02F5W CNS02F5W
C 24	37.6	3.3	1402	10	BG395673 BG395673
C 25	37.4	3.3	639	10	BH658123 BH658123
C 26	37.2	3.3	1201	13	BX381961 BX381961
C 27	36.8	3.2	636	12	BG820861 BG820861
C 28	36.6	3.2	242	9	AL437648 AL437648
C 29	36.6	3.2	485	10	BG650579 BG650579
C 30	36.6	3.2	595	14	CA785996 CA785996
C 31	36.6	3.2	639	10	BG415045 BG415045
C 32	36.6	3.2	1101	29	CNS016LW CNS016LW
C 33	36.4	3.2	692	12	BM596227 BM596227
C 34	36.2	3.2	374	12	BT780520 BT780520
C 35	36.2	3.2	528	14	CA029424 CA029424
C 36	36.2	3.2	628	13	BQ739865 BQ739865
C 37	36	3.2	334	9	AW508862 AW508862
C 38	36	3.2	925	29	CNS0091P CNS0091P
C 39	36	3.2	1201	13	BX444391 BX444391
C 40	35.8	3.1	292	10	BR431085 BR431085
C 41	35.8	3.1	645	29	CNS012T3 CNS012T3
C 42	35.8	3.1	1163	9	AL577643 AL577643
C 43	35.6	3.1	317	14	CB243032 CB243032
C 44	35.6	3.1	637	10	BS581130 BS581130
C 45	35.6	3.1	849	13	BX462111 BX462111

ALIGNMENTS

RESULT 1
BF662931
LOCUS
DEFINITION 963039P08.y1 C. reinhardtii CC-1690. Stress condition I, normalized
, lambda Zap II Chlamydomonas reinhardtii cDNA, mRNA sequence.

ACCESSION BF662931.1 GI:12253075
VERSION
KEYWORDS
SOURCE
ORGANISM Chlamydomonas reinhardtii
Chlamydomonas reinhardtii
Eukaryota, Viridiplantae, Chlorophyta; Chlorophyceae; Volvocales;
Chlamydomonadales; Chlamydomonadales.

REFERENCE
AUTHORS
TITLE
Lefebvre, P., McDermott, J.P., Shirger, J., Sillfow, C. and Stern, D.
Analyses of the Chlamydomonas reinhardtii Genome: A Model,
Unicellular System for Analyzing Gene Function and Regulation in
Vascular Plants; project phase 3

JOURNAL
COMMENT
Unpublished
Contact: Charles Hauser
DCMB Box 91000
Duke University
Durham, NC 27708-1000
Tel: 919 613 8159
Fax: 919 613 8177
Email: chauser@duke.edu
Location/Qualifiers

FEATURES

source

1. 767
 /organism="Chlamydomonas reinhardtii"
 /mol_type="mRNA"
 /strain="CC-1690 wild type mt+ 21gr"
 /db_xref="taxon:3055"
 /clone_1lb="C. reinhardtii CC-1690, stress condition 1, normalized, lambda Zap II"
 /note="Vector: pBluescript II SK-; Site 1: EcoRI; Site 2: XhoI; This library, constructed by John Davies and Jeffrey McDermott, combines cDNAs from CC-1690 cells grown to mid-log phase in YAP-N (30 min, 1hr, 4hr), YAP-S (30 min, 1hr, 4hr), YAP-P (4hr, 12hr, 24hr), NO3 (30min, 1hr, 4hr) and NH4 to NO3 (30min, 1hr, 4hr). PolyA mRNA was purified from each sample, pooled and cDNA synthesized. The cDNA was directionally cloned into lambda Zap II (Stratagene) in the EcoRI (5') and XhoI (3') sites. pBluescript II SK- plasmids were excised from the lambda Zap clones by superinfection with ExAssist (Stratagene) phage. The library was normalized using method 4 described in Bonaldo et al (1996) Genome Research 6: 791-806."

BASE COUNT 133 a 238 c 288 g 107 t 1 others

ORIGIN

Query Match 3.8%; Score 43; DB 10; Length 767;
 Best Local Similarity 48.6%; Pred. No. 0.26;
 Matches 118; Conservative 0; Mismatches 125; Indels 0; Gaps 0;

QY 390 GGCAGCTTGGAAAGCTGCGACCAACGGTCTGTTGGCCATGAGAACTGAAATGACGG 449
 DB 190 GGCAGCTTGGAAAGCTGCGACCAACGGTCTGTTGGCCATGAGAACTGAAATGACGG 249
 QY 450 CTGCGACGAGTACCTGCGCGCTGTGTAATCACTCAACGAAAGAGCGCTGACATGCA 509
 DB 250 CGGAGAGGACTACTCGCTGCGAATACCGCGCTGAGAGAGACGTGTCAAGAGGCGCA 309
 QY 510 CGTTACGCGCTACTACTGTTCACTGCTGAGTGAAGACATCTCTGAAATCTCTTC 569
 DB 310 CGTAAAGGTGAAGCGCAAGCGGTGGCGCGCGCGCAACCTCTTCCGCGCTGTGGA 369
 QY 570 GGATGCGATGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 629
 DB 370 GCGCGCGGTGCG 429

QY 630 ACG 632
 DB 430 GCG 432

RESULT 2

CC304102

LOCUS CH261-22111 Sp6.1 CH261 Gallus gallus genomic clone CH261-22111,
 DEFINITION genomic survey sequence.
 ACCESSION CC304102
 VERSION CC304102.1 GI:30675543
 KEYWORDS GGS.
 ORGANISM Gallus gallus (chicken)
 SOURCE
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
 Phasianidae; Gallus.
 1 (bases 1 to 1071)
 AUTHORS Kramitzki, C., Higginbotham, J., Wylie, K., Carter, J., McPherson, J.,
 Warren, W., Graves, T., Mardis, E. and Wilson, R.
 TITLE Gallus gallus BAC End Reads
 JOURNAL Unpublished
 COMMENT Contact: Richard K. Wilson
 Genome Sequencing Center
 Washington University School of Medicine
 Email: submissions@wustl.edu
 Insert Length: 182000 Std Error: 0.00
 Seq primer: Sp6 ATTAGGTGACACTATAG
 Class: BAC ends

High quality sequence start: 27
 High quality sequence stop: 476.
 Location/Qualifiers
 1. 1071

FEATURES

source

/organism="Gallus gallus"
 /mol_type="genomic DNA"
 /strain="Red Jungle Fowl"
 /db_xref="taxon:9031"
 /clone="CH261-22111"
 /sex="female"
 /cell_line="UCD001, inbred 256"
 /clone_1lb="CH261"
 /note="Vector: pTARHAC 1; Site 1: EcoRI; Site 2: EcoRI;
 CH261 Female Chicken library - for library and clone
 ordering information: http://www.chori.org/bacpac"

BASE COUNT 254 a 312 c 213 g 290 t 2 others

ORIGIN

Query Match 3.8%; Score 43; DB 29; Length 1071;
 Best Local Similarity 53.2%; Pred. No. 0.31; 80; Indels 0; Gaps 0;
 Matches 91; Conservative 0; Mismatches 80; Indels 0; Gaps 0;

QY 840 TGAATTGTTGATGAGAGCGTGGCAATCGGTGTCCTCGGATTTGGTCCGAGC 899
 DB 69 TGTGTGCTGCGCGGCGCTGCGACAGGTGAGTTGACCCGATGTGTGCCACGTG 128
 QY 900 TGTCTTGGGCGCAATACAGATGCTGATGCTGAGAGAGAAATCTCCCGGT 959
 DB 129 TGGACATGATTAACATGCTGATGCTGCTGCTGCGAGAGAAAGCTTAACCTGCA 188
 QY 960 CAGGTTGCGATCACTTCGCGCGCGATCGGATGATGATGATGATGATGATGATGATGAT 1010
 DB 189 CCGGTTCTGTATACAGCTGTCTGACTCTTGTGAGAGGTGAGCTGACGTGCTGC 239

RESULT 3

CD432432

LOCUS ETH1_30_H03_b1_A002 Ecthylene-treated seedlings Sorghum bicolor cDNA
 DEFINITION
 CD432432
 ACCESSION CD432432.1 GI:31348075
 VERSION CD432432.1 GI:31348075
 KEYWORDS EST.
 ORGANISM Sorghum bicolor (sorghum)
 SOURCE
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
 clade; Panicoideae; Andropogoneae; Sorghum.
 1 (bases 1 to 609)
 AUTHORS Cordomier-Pratt, M.-M., Wentzel, V., Suzuki, Y., Sugano, S., Klein,
 R.R., Liang, C., Sun, F., Sullivan, R., Shah, M., Salzman, R., Chua Tan,
 N., Gonzalez, M., Lane, S., Miller, V., Nanda, P., Olaseinde, O.,
 Baetman, A. and Pratt, L.H.
 TITLE An EST database from Sorghum: 1-aminocyclopropane-1-carboxylic acid
 (ACC)-treated seedlings
 JOURNAL Unpublished
 COMMENT Other ESTs: ETH1_30_H03_G1_A002
 Contact: Cordomier-Pratt MM
 Laboratory for Genomics and Bioinformatics
 The University of Georgia, Department of Plant Biology
 Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
 Tel: 706 542 1860
 Fax: 706 583 0210
 Email: mmpratt@uga.edu

Library constructed by Dr. Yutaka Suzuki and Dr. Sumio Sugano in
 the Human Genome Center, University of Tokyo Institute of Medical
 Science; plant material and RNA prepared at Texas A & M University;
 sequencing done in the laboratory for Genomics and Bioinformatics,
 University of Georgia. Sequence ends have been trimmed to exclude
 vector and regions below Phred quality 16. Three-prime sequences
 are presented as their reverse complement and have been trimmed to
 exclude polyA.
 Seq primer: Sug3 (CGACCTGACCTCGAGACA)

Unpublished
Contact: Genoscope

Genoscope 3 Centre National de Séquençage
BP 191 91006 Evry cedex - France
Email: segr@genoscope.cns.fr, Web: www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of
Invitrogen. This sequence belongs to sequence cluster 9540.f For
more information about this cluster, see
<http://www.genoscope.cns.fr/cgi-bin/cluster.cgi?seqid=CSDDC01B5B100P1&cluster=9540.f>. Contact :
Peng Liang Email: liang@lifetech.com URL: <http://fulllength.invitrogen.com/> Invitrogen Corporation 1600
Paradise Avenue Genoscope, sequence ID : CSDDC01B5B100P1.

```

/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CSDDC018Y020"
/tissue_type="NEUROBLASTOMA COT 25-NORMALIZED"
/clone_id="Homo sapiens NEUROBLASTOMA COT 25-NORMALIZED"
/note="1st strand cDNA was primed with a NotI-Cligo (47)
primer. Five prime end enriched, double-strand cDNA was
digested with Not I and cloned into the Not I and EcoR V
sites of the pCMVSPORT 6 vector. Library was normalized."
BASE COUNT      192 a      357 c      358 g      218 t
ORIGIN

```

ch	Score	DB	Length
1 Similarity	37.6%	Pred. No. 1.5;	
70; Conservative	34;	Mismatches	82; Indels 0; Gaps 0

Best Local Similarity	37.08%	Friedl, NC, 1.5/							
Matches	70;	Conservative	34;	Mismatches	82;	Indels	0;	Gaps	0

DY 782 AATTCGCTGTGATGCAGTGGCGGGGATCCAAACGTTGGAACGTGGCAGAATTGG 841
| | : : | | : | : | : | : : ||||
Db 1153 AAGAYSCGATTGCTGTGGGAAGKGTGCCCAACCCKTKGWGCCMMCGAATAKGDS 109

DY 842 AGTTTGTCGATGGGACCGTCCGGCAATCGCGTGCCTCCGTGAAATTGCCGTGCCAGCTG 901
| : : | : | : : : : : : : : : : : : : : : :
Db 1093 CCCCTTCGYBGTGCCCCCTKGCYCSCCCCAGCAKKGYGKKGAATTKSSCAACCGCAGCTK 103

DY
902 GTCTTGGGCGAGAACTACAGATGCTCAGATCGTTGAGCAGAGAATCTGCCCCGTCA 961

DH
1033 GSCATTCGAATGCCTGATCATATTCTTGACAATCVASCTTAASWAGTAGGCCCCGATTTGGGGAGACTGATCC 974

0y 962 TGETTG 967
||| :|

RESULT 6

BI440943/c	BI440943	556 bp	mRNA	linear	EST 12-MAR-2001
LOCUS					
DEFINITION	ic78e08.y1 Melton Normalized Mixed Mouse Pancreas 1 NI-MMS1 Mus				

ACCESSION
VERSION
B1440943.1
GI:15265633

KEYWORDS	EST.
SOURCE	Mus musculus (house mouse)
ORGANISM	Mus musculus
	Eukaryota : Metazoa : Chordata : Mammalia : Theriobata : Eutlaestromia :

REFERENCE
AUTHORS
1 (bases 1 to 556)
Melton, D., Brown, J., Kenty, G., Permutt, A., Lee, C., Kaestner, K.,
Zundayola, Melazoa;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 556)
Melton, D., Brown, J., Kenty, G., Permutt, A., Lee, C., Kaestner, K.,

Lemishka, I., Scarce, M., Brestelli, J., Gradwohl, G., Clifton, S., Hillier, L., Marra, M., Pape, D., Wylie, T., Martin, J., Blaisnail, A., Schmitt, A., Theising, B., Ritter, E., Ronko, I., Bennett, J., Cardenas M., Gibbons M., McCann R., Cole P., Tsacoreishvili R., Williams T.

1980
1981
1982
1983
1984
1985
1986
1987
1988
1989
1990
1991
1992
1993
1994
1995
1996
1997
1998
1999
2000
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
2020
2021
2022
2023
2024
2025
2026
2027
2028
2029
2030
2031
2032
2033
2034
2035
2036
2037
2038
2039
2040
2041
2042
2043
2044
2045
2046
2047
2048
2049
2050
2051
2052
2053
2054
2055
2056
2057
2058
2059
2060
2061
2062
2063
2064
2065
2066
2067
2068
2069
2070
2071
2072
2073
2074
2075
2076
2077
2078
2079
2080
2081
2082
2083
2084
2085
2086
2087
2088
2089
2090
2091
2092
2093
2094
2095
2096
2097
2098
2099
2100
2101
2102
2103
2104
2105
2106
2107
2108
2109
2110
2111
2112
2113
2114
2115
2116
2117
2118
2119
2120
2121
2122
2123
2124
2125
2126
2127
2128
2129
2130
2131
2132
2133
2134
2135
2136
2137
2138
2139
2140
2141
2142
2143
2144
2145
2146
2147
2148
2149
2150
2151
2152
2153
2154
2155
2156
2157
2158
2159
2160
2161
2162
2163
2164
2165
2166
2167
2168
2169
2170
2171
2172
2173
2174
2175
2176
2177
2178
2179
2180
2181
2182
2183
2184
2185
2186
2187
2188
2189
2190
2191
2192
2193
2194
2195
2196
2197
2198
2199
2200
2201
2202
2203
2204
2205
2206
2207
2208
2209
2210
2211
2212
2213
2214
2215
2216
2217
2218
2219
2220
2221
2222
2223
2224
2225
2226
2227
2228
2229
2230
2231
2232
2233
2234
2235
2236
2237
2238
2239
2240
2241
2242
2243
2244
2245
2246
2247
2248
2249
2250
2251
2252
2253
2254
2255
2256
2257
2258
2259
2260
2261
2262
2263
2264
2265
2266
2267
2268
2269
2270
2271
2272
2273
2274
2275
2276
2277
2278
2279
2280
2281
2282
2283
2284
2285
2286
2287
2288
2289
2290
2291
2292
2293
2294
2295
2296
2297
2298
2299
2300
2301
2302
2303
2304
2305
2306
2307
2308
2309
2310
2311
2312
2313
2314
2315
2316
2317
2318
2319
2320
2321
2322
2323
2324
2325
2326
2327
2328
2329
2330
2331
2332
2333
2334
2335
2336
2337
2338
2339
2340
2341
2342
2343
2344
2345
2346
2347
2348
2349
2350
2351
2352
2353
2354
2355
2356
2357
2358
2359
2360
2361
2362
2363
2364
2365
2366
2367
2368
2369
2370
2371
2372
2373
2374
2375
2376
2377
2378
2379
2380
2381
2382
2383
2384
2385
2386
2387
2388
2389
2390
2391
2392
2393
2394
2395
2396
2397
2398
2399
2400
2401
2402
2403
2404
2405
2406
2407
2408
2409
2410
2411
2412
2413
2414
2415
2416
2417
2418
2419
2420
2421
2422
2423
2424
2425
2426
2427
2428
2429
2430
2431
2432
2433
2434
2435
2436
2437
2438
2439
2440
2441
2442
2443
2444
2445
2446
2447
2448
2449
2450
2451
2452
2453
2454
2455
2456
2457
2458
2459
2460
2461
2462
2463
2464
2465
2466
2467
2468
2469
2470
2471
2472
2473
2474
2475
2476
2477
2478
2479
2480
2481
2482
2483
2484
2485
2486
2487
2488
2489
2490
2491
2492
2493
2494
2495
2496
2497
2498
2499
2500
2501
2502
2503
2504
2505
2506
2507
2508
2509
2510
2511
2512
2513
2514
2515
2516
2517
2518
2519
2520
2521
2522
2523
2524
2525
2526
2527
2528
2529
2530
2531
2532
2533
2534
2535
2536
2537
2538
2539
2540
2541
2542
2543
2544
2545
2546
2547
2548
2549
2550
2551
2552
2553
2554
2555
2556
2557
2558
2559
2560
2561
2562
2563
2564
2565
2566
2567
2568
2569
2570
2571
2572
2573
2574
2575
2576
2577
2578
2579
2580
2581
2582
2583
2584
2585
2586
2587
2588
2589
2590
2591
2592
2593
2594
2595
2596
2597
2598
2599
2600
2601
2602
2603
2604
2605
2606
2607
2608
2609
2610
2611
2612
2613
2614
2615
2616
2617
2618
2619
2620
2621
2622
2623
2624
2625
2626
2627
2628
2629
2630
2631
2632
2633
2634
2635
2636
2637
2638
2639
2640
2641
2642
2643
2644
2645
2646
2647
2648
2649
2650
2651
2652
2653
2654
2655
2656
2657
2658
2659
2660
2661
26

COMMENT: Other_ESTs: 1c78e08_xl
Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue

Endocrine Pancreas Consortium
Harvard University, Howard Hughes Medical Institute
Dept of Molecular and Cellular Biology, 7 Divinity Ave,
MA 02138
Tel.: 617-495-1917

Per: 617-495-8612
 Fax: 617-495-8657
 Email: dmeitton@biohp.harvard.edu
 Library was constructed by Dr. Douglas Melton DNA sequencing by:
 Washington University Genome Sequencing Center For information on
 obtaining a clone please contact: Juliana Brown
 (brown@fas.harvard.edu)
 Possible reversed clone: similarity on wrong strand
 MG11946189 This sequence now available from the IMAGE consortium
 for clone orders contact: info@image.llnl.gov
 High quality sequence stop: 437.

```

/organism="Mus musculus"
/mol_type="mRNA"
/strain="ICR"
/db_xref="taxon:10090"
/clone="IMAGE:5659863"
/sex="both for embryonic & newborn, male for adult and
adult islet"
/dev_stage="Embryonic day 10.5, E12.5, E16.5, newborn,
adult, mixed"
/lab_host="MDH108"
/clone_lib="Wetlon Normalized Mixed Mouse Pancreas 1
11-MW51"

```

libraries representing E10.5/12.5 pancreatic bud, E16.5 pancreas, newborn pancreas, adult pancreas, and adult islets of Langerhans were separately constructed using

SuperScript Plasmid Library kit (Life Technologies). cDNA was made by oligo-dT priming and size-selected by column fractionation. Libraries were amplified once on solid

support and plasmid DNA from each library was prepared and mixed in equal amounts. The mixed library DNA was normalized by method #4 from Bonaldo, Lennon, and Soares 1996 Genome Research 6:791-806; 0.5 microgram

single-stranded mixed library plasmid DNA was mixed with 5 micrograms PCR product representing mixed library inserts and hybridized to an EcoT of 6. Single-stranded (unhybridized) plasmids were isolated by hydroxyapatite

chromatography and used to make this library."

3.6%; Score 40.6; DB 12; Length 556;
 48.5%; Pred. No. 1.1;
 0: Mismatches 119; Indels 0; Gaps 0

GAAATTCGATCTGCAGTTGGTATTGGAGTTGGTAGTGCGACGGAAAAAGTGG 755

GTGGCAACGTCGACCCCTGGGAATCGCTTGATGCAGTGGGGGGGGAATCACA 815
GAAAGCTGCCAACGCGCTGCTGGGCGAAGCTGGGGGCGCTGGGGGGCGCATGGCGA 4/8

GGAACTGTGGCGAGAATTTCAGTTTGGTTCGATGGGACGTCGGCAATTCGCGTG 875
GCGCAGCCCGGTGTGGCCCCCGACAGCGCCAGCGCTGTGCGGGCGCTGGAGCGCGTGC 418

TTGGA TTGCGCTGCCCCGAGACTGTTCTTTGGGGCAGAACTAACAGATGC 926
AAGAGCCCGTGAACCGGAGCCCTCAGGCGTGGCGCCGCTGGAGGAGACGGCGAGCGCGGC 358

307
 TAGGCGACAGTGCCTGCTGACCTTAGCGCGCTGTCTGGAGGAATTCGCGAC

737 bp mRNA linear EST 13-MAR-2002

-brw-a-11-0-UI.r1 NIH_BMAP_M_S4 Mus musculus cDNA clone

Best Local Similarity 48.5%; Pred. No. 1.3; Mismatches 119; Indels 0; Gaps 0;
Matches 112; Conservative 0;

QY 696 TCACGAGAAATGCACTGCGCTGTATGAGAGTTGATGCGCAGAGAAAGTGC 755
DB 304 TCCGTGAGAGCTGCAAGCGGCTGCGGCGAGAGCTGGGGCGGCTGGGGCGGATGCCA 363
QY 756 TCACATGAGCAACCGTGCACCTGTGAAATCGCTGTATGACATGAGGCGGATCCACA 815
DB 364 GGCCTGCGCAGCGGTGCGCCCGCAGACGCCAGCTGGTGGCGGCGCTGGAGCCGCTGC 423
QY 816 AGCGTTGAACTGTGCGCGAATTTGATTTGTTGATGATGAGGACGTCGCGGCAATCGCGNG 875
DB 424 TGCAGAGAGCGCTGACGCGAGCTTACGCTGCGCGCTGGAGAGACGCGAGCGCGGC 483
QY 876 GTCCCGTGAATTGCTGCGCCGAGCTGTCTTGGGAGCAACTAACAGATGC 926
DB 484 GACCCGAGCGACAGTGCCTGCGCTGAGCGCGCTGTGAGAGAACTGCGGC 534

RESULT 9
BU591391/c 937 bp mRNA linear EST 07-JAN-2003
LOCUS AGENCOURT_8783384 NIH_MGC_137 Mus musculus cDNA clone IMAGE:6431460
DEFINITION 5', mRNA sequence.
ACCESSION BU591391
VERSION BU591391.2 GI:27538019
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 937)
NHL-MGC http://mgc.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished
On Sep 19, 2002 this sequence version replaced gi:23243003.
Contact: Robert Strausberg, Ph.D.
Email: cgaabs-r@mail.nih.gov
Tissue Procurement: Gerard Gradwohl (PNAS 97 P1607-1611, 2000)
cDNA Library Preparation: Catherine Lee, Endocrine Pancreas Consortium
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at:
http://image.lnl.gov
Plate: IRBD16 row: f column: 01
High quality sequence stop: 660.

FEATURES
source
Location/Qualifiers
1..937
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="IMAGE:6431460"
/lab_host="DH10B"
/clone_id="NIH_MGC_137"
/note="Organ: pancreas; Vector: pSPORT1; site: 1; Salt; Site 2: NCI; Library consists of a pool of clones rearrayed from the following libraries: Melton normalized mixed mouse pancreas 1 NI-MMS1, Amplified Melton mouse islets 1 MSL-A, and Kaestner ngn3 wt. Clones rearrayed in the laboratory of K. Kaestner (University of Pennsylvania). Note: this is a NIH MGC Library."

BASE COUNT 183 a 293 c 249 g 212 t
ORIGIN
Query Match 3.6%; Score 40.6; DB 13; Length 937;
Best Local Similarity 48.5%; Pred. No. 1.5;
Matches 112; Conservative 0; Mismatches 119; Indels 0; Gaps 0;

QY 696 TCACGAGAAATGCACTGCGCTGTATGAGAGTTGATGCGCAGAGAAAGTGC 755
DB 304 TCCGTGAGAGCTGCAAGCGGCTGCGGCGAGAGCTGGGGCGGCTGGGGCGGATGCCA 363
QY 756 TCACATGAGCAACCGTGCACCTGTGAAATCGCTGTATGACATGAGGCGGATCCACA 815
DB 364 GGCCTGCGCAGCGGTGCGCCCGCAGACGCCAGCTGGTGGCGGCGCTGGAGCCGCTGC 423
QY 816 AGCGTTGAACTGTGCGCGAATTTGATTTGTTGATGATGAGGACGTCGCGGCAATCGCGNG 875
DB 424 TGCAGAGAGCGCTGACGCGAGCTTACGCTGCGCGCTGGAGAGACGCGAGCGCGGC 483
QY 876 GTCCCGTGAATTGCTGCGCCGAGCTGTCTTGGGAGCAACTAACAGATGC 926
DB 484 GACCCGAGCGACAGTGCCTGCGCTGAGCGCGCTGTGAGAGAACTGCGGC 534

DB 562 TCCGTGAGAGAGCTGACGCGCTGCGGCGAGAGCTGCGGCGGCTGCGCGAGATGCCA 503
QY 756 TCACATGAGCAACCGTGCACCTGTGAAATGCTGTTGATGATGAGTGCAGAGATCCACA 815
DB 502 GGCCTGCGCAGCGGTGCGCCCGCAGACGCCAGCTGGTGGCGGCGCTGGAGCGCGCTGC 443
QY 816 AGCGTTGAACTGTGCGCGAATTTGATTTGTTGATGATGAGGACGTCGCGGCAATCGCGNG 875
DB 442 TGCAGAGAGCGCTGACGCGAGCTTACGCTGCGCGCTGGAGAGACGCGAGCGCGGC 383
QY 876 GTCCCGTGAATTGCTGCGCCGAGCTGTCTTGGGAGCAACTAACAGATGC 926
DB 382 GACCCGAGCGACAGTGCCTGCGCTGAGCGCGCTGTGAGAGAACTGCGGC 332

RESULT 10
AK084367/c 1122 bp mRNA linear HTC 05-DEC-2002
LOCUS Mus musculus 12 days embryo eyeball cDNA, RIKEN full-length enriched library, clone:D230032H18 product: pentaxin related gene, full insert sequence.
DEFINITION AK084367
ACCESSION AK084367
VERSION AK084367.1 GI:26101923
KEYWORDS HTC; CAP trapper.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1
Carninci, P. and Hayashizaki, Y.
High-efficiency full-length cDNA cloning
Meth. Enzymol. 303, 19-44 (1999)
99279253
10349636
2
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Komno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
Genome Res. 10 (10), 1617-1630 (2000)
11042159

REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
20499374
11042159
3
Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Komno, H., Akiyama, J., Nishi, K., Katsunai, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishino, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
RIKEN integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multicapillary sequencer
Genome Res. 10 (11), 1757-1771 (2000)

REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
20530913
11076861
4
Kawai, J., Shinagawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y., Aizawa, K., Hara, A., Fukunishi, Y., Komno, H., Adachi, J., Fukuda, S., Aizawa, K., Izawa, M., Nishi, K., Kiyosawa, H., Kondo, S., Yamada, I., Kado, T., Okazaki, Y., Gotohori, T., Bono, H., Kasukawa, T., Saito, R., Kadota, K., Matsuda, H., Ashburner, M., Batalov, S., Casavant, T., Pleischmann, W., Gaasterland, T., Gissi, C., King, B., Kochiwa, H., Koehn, P., Lewis, S., Matsuo, Y., Nikaido, I., Pesole, G., Tomita, M., Quackenbush, J., Schriml, J. M., Staudil, F., Suzuki, R., Tomita, M., Wagner, L., Washio, T., Sakai, K., Okido, T., Furuno, M., Aono, H., Balderelli, R., Barsh, G., Blake, J., Boffelli, D., Bojunga, N., Carninci, P., de Bonaldo, M. F., Brownstein, M. J., Bult, C., Fletcher, C., Fujita, M., Gariboldi, M., Gustincich, S., Hill, D., Hofmann, M., Hume, D. A., Kamiya, M., Lee, N. H., Lyons, P., Marchionni, L., Mashima, J., Mazzarelli, J., Mombaeys, P., Nordone, P., Rung, B., Ringwald, M., Rodriguez, I., Sakamoto, N., Sasaki, H., Sato, K., Schonbach, C., Seya, T., Shibata, Y., Storch, K. F., Suzuki, H., Toyooka, K., Wang, K. H., Weitz, C., Whitaker, C., Wilmberg, L.,

402 AGCTGCAACCAAGTCTGTTGCGC 427

D0 26 AGCAGCTTGAAGTTCATTGAGGAGAGGAGATTTGACACAGGAGATTCAGAGGTGGATTCCG 202
 D1 741 CGACGAAAAGATGTGTGCACATGGCGAACCGTGCAACCTGGGAAATGCTGTGATCAAGT 800
 D2 201 AGACGAGGAGATGGAGAGAGATGTTCTCCGGGCACTGGAGATGTTGAGCATGAGATGCA 142

Qy 801 GGGCGGGATCCACAGCGTGAAC 827
|||
Db 141 CGCGCGGATGAGAGCCGCGAGGTA 115
|||

Search completed: August 22, 2003, 10:35:38
Job time : 1962 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using fixme_plus.p2n model

Run on: August 22, 2003, 09:13:28 ; Search time 3108 Seconds
(without alignments)
4988.658 Million cell updates/sec

Title: US-10-007-527a-2
Perfect score: 2005
Sequence: 1 MTSVSAEHLGKDRPVLVS.....HLHYRLPADVPRPPIISVRK 379

Scoring table: BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Zgapop 6.0, Zgapext 7.0
Delop 6.0, Delext 7.0

Searched: 2888711 seqs, 2045481386 residues

Total number of hits satisfying chosen parameters: 5777422
Minimum DB seg length: 0
Maximum DB seg length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODE=frame+ p2n.model -DEV=xlh
-O=/cgn2_1/USPD00007527/funat 20082003 165151 15366/app query.fasta 1.519
-DB=Genembl -GEMT=fastcap -SUFFIX=p2n.rge -MINMATCH=0.1 -LOOPTCT=0 -LOOPTXT=0
-INITs-bits -STRPT=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALLIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US10007527@cgn 1.1 2372 @funat 20082003 165151 15366 -NCPU=6 -ICPU=3
-NO WMAP -LARGEBLOCK -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : GenBml.*
1: gb_da.*
2: gb_hcg.*
3: gb_in.*
4: gb_om.*
5: gb_ov.*
6: gb_pat.*
7: gb_ph.*
8: gb_pl.*
9: gb_pr.*
10: gb_ro.*
11: gb_sts.*
12: gb_sy.*
13: gb_un.*
14: gb_vl.*
15: em_da.*
16: em_fun.*
17: em_hum.*
18: em_in.*
19: em_mu.*
20: em_om.*
21: em_or.*
22: em_ov.*
23: em_pat.*
24: em_ph.*
25: em_pl.*
26: em_ro.*
27: em_sts.*
28: em_un.*

29: em_vl.*
30: em_hcg_hum.*
31: em_hcg_inv.*
32: em_hcg_other.*
33: em_hcg_mus.*
34: em_hcg_pln.*
35: em_hcg_rtd.*
36: em_hcg_vrt.*
37: em_hcg_vrt.*
38: em_sy.*
39: em_hcg_hum.*
40: em_hcg_mus.*
41: em_hcg_other.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2005	100.0	1140	AX548642	AX548642 Sequence
2	2005	100.0	6334	AY178757	AY178757 Rhodococ
3	2005	100.0	6334	AX548646	AX548646 Sequence
4	2005	100.0	9652	AX548648	AX548648 Sequence
5	2005	100.0	11241	AX548647	AX548647 Sequence
6	2005	100.0	11241	AY180162	AY180162 Shuttle v
7	609.5	30.4	2439	APU83788	U83788 Actinomyc
8	569	28.4	3540	AY150274	AY150274 Propionib
9	497	24.8	2051	FFR662	FFR662 Propionib
10	453	22.6	2661	SLMG33A	SLMG33A Propionib
11	452.5	22.6	9367	SNA243257	SNA243257 S.lavendula
12	425.5	21.2	4603	AY172684	AY172684 Coryneb
13	399.5	19.9	2297	E17316	E17316 Bifidobacte
14	392.5	19.6	5250	AF085719	AF085719 Bifidobac
15	374.5	18.7	5804	AF428260	AF428260 Propionib
16	360.5	18.0	12855	AF533985	AF533985 Streptomy
17	346.5	17.3	10992	STMPN22	STMPN22 Streptomy
18	345	17.2	3748	AB032347	AB032347 Cloning v
19	345	17.2	5128	AB084167	AB084167 Pseudomon
20	345	17.2	5128	CCA308231	CCA308231 Coryneb
21	342.5	17.1	4109	AY072037	AY072037 Expressio
22	337	16.8	5349	AY072037	AY072037 Expressio
23	337	16.8	6342	AY072040	AY072040 Expressio
24	337	16.8	6607	AY072039	AY072039 Expressio
25	337	16.8	8058	AY072038	AY072038 Expressio
26	337	16.8	8215	P144669	P144669 Cloning v
27	337	16.8	8830	P144669	P144669 Cloning v
28	333.5	16.6	2297	AF273219	AF273219 Pseudomon
29	326	16.3	3147	AF092035	AF092035 Integrati
30	326	16.3	4447	AF092037	AF092037 Coryneb
31	326	16.3	4447	AR173114	AR173114 Sequence
32	326	16.3	4447	BD106940	BD106940 Method fo
33	326	16.3	4447	BD139585	BD139585 Arginine
34	326	16.3	4457	B1PLASMI	B1PLASMI Brevibacter
35	326	16.3	5135	SVE133193	SVE133193 Shuttle v
36	326	16.3	5216	SVE133194	SVE133194 Shuttle v
37	326	16.3	6241	AF092036	AF092036 Shuttle v
38	326	16.3	6272	AF091268	AF091268 Shuttle v
39	326	16.3	6398	AF092931	AF092931 Shuttle v
40	326	16.3	6601	SVE133195	SVE133195 Shuttle v
41	318	15.9	4447	AR173115	AR173115 Sequence
42	318	15.9	4447	BD106941	BD106941 Method fo
43	318	15.9	4447	BD139586	BD139586 Arginine
44	315	15.7	7212	AF327371	AF327371 Streptococ
45	296	14.8	4448	COGAM30	COGAM30 Corynebacte

RESULT 1

ALIGNMENTS

AX548642
 LOCUS AX548642 1140 bp DNA linear PAT 27-NOV-2002
 DEFINITION Sequence 1 from Patent WO2055709.
 ACCESSION AX548642
 VERSION AX548642.1 GI:25813612
 KEYWORDS
 SOURCE
 ORGANISM
 Rhodococcus erythropolis
 Rhodococcus erythropolis
 Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 Corynebacterineae; Nocardiaceae; Rhodococcus.
 REFERENCE
 1
 Brummel, M.G., Cheng, Q., Kostichka, K.N. and Tomb, J.F.
 Rhodococcus cloning and expression vectors
 Patent: WO 02055709-A1 18-Jul-2002
 JOURNAL
 B.I. DU PONT DE NEMOURS AND COMPANY (US)
 Location/Qualifiers
 source
 1..1140
 /organism="Rhodococcus erythropolis"
 /mol_type="genomic DNA"
 /strain="AN12"
 /db_xref="taxon:1833"
 BASE COUNT 241 a 287 c 347 g 265 t
 ORIGIN
 Alignment Scores:
 Pred. No.: 7.8e-119 Length: 1140
 Score: 2005.00 Matches: 379
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 6 Gaps: 0
 US-10-007-527a-2 (1-379) x AX548642 (1-1140)
 QY 1 MetThrSerValSerAlaGluHisLeuSerGlyLysAspArgProProValLeuValSer 20
 Db 1 ATGACCAAGGTTGAGTTCGAGGAGGCTTCGCTCGGAGAAAGCCGCTCCGCTCTGTCG 60
 QY 21 SerAspLysArgGlyIleArgHisGluLeuArgProLysLeuGlnIleLeuThrSer 40
 Db 61 TCCGTAAGCCGGGATCCGAGACATCGACACCAACTTCAAAATACACACCTCA 120
 QY 41 GluThrPheAsnAlaCysGlyArgProLysSerGlyValaGlnGlyValThrIleValAsn 60
 Db 121 GAAACATTTAAGCCTGTGCGCGCGGATTTCTGCGGAGACGCGTGTGACCATTTGTAAC 180
 QY 61 GlyProLysGlySerGlyPheGlyGlyLeuArgSerCysGlyLysGlyTyrIleCysPro 80
 Db 181 GGTCCGAAAGGTTCTGAGATTCGAGGCTTCGCTCGGAGAAAGGCTGAGATCTGCC 240
 QY 81 CysCysAlaGlyLysValGlyAlaHisArgAlaAspGlnIleSerGlnValValaHis 100
 Db 241 TCGTGTGCGGAGAAAGTGGTGCATCGTGCAGCAAAATTTCTCAAGTTGTTGCTCAT 300
 QY 101 GlnLeuGlyThrGlySerValAlaMetValThrMetThrMetAlaGHisThrAlaGlyGln 120
 Db 301 CAACCTCGGAGCTGATCTGTTGCGATGTGAGAGATCAAGCCGATACAGCTGTGAC 360
 QY 121 ArgLeuHisAspLeuThrThrGlyLeuSerAlaAlaTyrPlyAlaAlaThrAsnGlyArg 140
 Db 361 CGGCTCCAGACCTTGTGACTTTCGGAGCTCGAAGCTGGAAGCTGGACCAAGCTGCT 420
 QY 141 ArgTyrArgThrGluArgGluMetTyrGlyCysAspGlyTyrValaArgAlaValaGluIle 160
 Db 421 CGTTGGCTTACCGAAGCTGTAATGTACGCTGCGAGATACGTCGCGCTGTTGAAATC 480
 QY 161 ThrHisGlyLysAsnGlyTyrHisValaHisValaHisAlaLeuLeuMetPheSerGlyAsp 180
 Db 481 ACTCAAGGAAAGAAAGCTGGGACCTGCACGTTTCACGCTACTCATGTTCTGATGAC 540
 QY 181 ValSerGluAsnIleLeuGluSerPheSerAspAlaMetPheAspArgTyrThrSerLys 200
 Db 541 GTGAGTGAAGAACATCTTCTGATCTTCTGATGCGATGCGATGCGATGCGATCTCCAAA 600

QY 201 LeuValSerLeuGlyPheAlaAlaProLeuArgPheSerGlyLysLeuAspValArgLys 220
 Db 601 CTGATATCTCGGAGTTTGTGCGGCACTACGTAATCGGCTGCTCGAGTACGAAAG 660
 QY 221 IleGlyGlyAlaAspGlnValLeuAlaAlaTyrLeuThrLysIleAlaSerGlyVal 240
 Db 661 ATCCGCGGTGAAGCTGATCAAGTCTCTCGCTCGATTCGACGAAATTTGATCTGGGCTT 720
 QY 241 GlyMetGluValGlySerGlyLysAspGlyLysSerGlyArgHisGlyAsnArgAlaProTyr 260
 Db 721 GGTATGAGGTTGTGATGCGGACCGGAAAGGTGTCAACATGGCAACCGTGCACCTCG 780
 QY 261 GlnIleAlaValaAspAlaValaGlyLysAspProGlnAlaLeuGluLeuThrArgGluPhe 280
 Db 781 GAAATCGCTGTGATGAGCTGGGCGGATTCACAGGTTGGAACCTGTGGCGAGAAATT 840
 QY 281 GluPheGlySerMetGlyArgArgAlaIleAlaTyrSerArgGlyLeuArgAlaArgAla 300
 Db 841 GATTTGGTTGATGAGAGCTGGGCGGATTCGCGGATTCGCGGATTCGCGGATTCGCGG 900
 QY 301 GlyLeuGlyAlaGluLeuThrAspAlaGlnIleValaGlnGlnGlnLysSerAlaProVal 320
 Db 901 GGTCTTGGGCGAAGCTAACAGATGCTCAGATCTGTGAGCAGAAATTCGCCCGGCTC 960
 QY 321 MetValaAlaIleIleProAlaArgSerTyrMetIleArgThrCysAlaProTyrVal 340
 Db 961 ATGCTTGGATCATTCGCGGCGGATTCGAGATGATGATGATGATGATGATGATGATGAT 1020
 QY 341 PheGlyGluIleLeuGlyLysValaGluAlaGlyAlaThrTyrPheGluAsnLeuArgAspHis 360
 Db 1021 TTCGGGAGATCTCGGACTCGTCCAGCTGGCGGACCTTGGGAAATCTTCGATCAC 1080
 QY 361 LeuHisTyrArgLeuProAlaAlaAspValaArgProProIleIleSerValArgLys 379
 Db 1081 TTGCATTTATCATTTGCTCCGCGACGAGATGTGCGGCCCCGATATATTCGTTGCAAG 1137
 RESULT 2
 AY178757/c 6334 bp DNA linear BCT 01-MAR-2003
 LOCUS AY178757
 DEFINITION Rhodococcus erythropolis cryptic plasmid pAN12, complete sequence.
 ACCESSION AY178757
 VERSION AY178757.1 GI:28628256
 KEYWORDS
 SOURCE
 ORGANISM
 Rhodococcus erythropolis
 Rhodococcus erythropolis
 Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 Corynebacterineae; Nocardiaceae; Rhodococcus.
 REFERENCE
 1
 Kostichka, K., Tao, L., Brummel, M., Tomb, J.-F., Nagarajan, V. and
 Cheng, Q.
 A small cryptic plasmid from Rhodococcus erythropolis:
 Characterization and utility for gene expression
 Unpublished
 Direct Submission
 Submitted (11-NOV-2002) CRD, E.I. Dupont de Nemours Inc.,
 Experimental Station, Wilmington, DE 19880-0328, USA
 TITLE
 JOURNAL
 AUTHORS
 REFERENCES
 FEATURES
 source
 1..6334
 /organism="Rhodococcus erythropolis"
 /mol_type="genomic DNA"
 /db_xref="taxon:1833"
 /plasmid="pAN12"
 /note="cryptic plasmid"
 /complement(1912..3051)
 /gene="rep"
 /complement(1912..3051)
 /gene="rep"
 /note="Rep"
 /codon_start=1

```

/ctranal_table=11
/product="repllication protein"
/protein_id="AA031602.1"
/db_xref="GI:28628257"
/translation="MTSVSAEHLGSKDRPPVLSVDKRGIRHLEAPKIQITSTSTEN
ACGRPISGVNGVITVANGPKSGFGJLRSCKGKMI CPCCGKATGARADEISGVVHOL
GTGSAVMTWTMTHTAGORLHDLMTGLSAAMKATIGKRWTEREMAGDGVRAVEI
THKNGHVAHVHALLMPSGVSNTLESNDAMPDRTSKLSLGPAAPLRNSGLDV
RKIGRADVDLAATLVTKIASGVMEVSGDGSKGRNAPMEIIVDAVCGDPOALEL
KREFRSGMRRAIRMSRGLRPARAGIATDQIVYQESAPVWVAIIPASMMNIR
TCAPVFEITLIGVEAGATWENLDHLHYRLPADVRPPIISVRK"
complement(4288..5178)
/gene="div"
/gene="div"
/feature="div"
/codon_start=1
/ctranal_table=11
/product="putative cell division protein"
/protein_id="AA031603.1"
/db_xref="GI:28628258"
/translation="MDQDTPIALIGMNLAPWVVDIAKDAHWLJGKTRSGKSGOC
TYRLAAGSNPAVRVYGVDPSTSVLAPVHRRPAPNIELELNDPKYLRLOFVKA
ESDRKIBCEMDRIKISLSPALPILVLESPGILGADQFPAATGKPKADRYAP
RITSLVQAIAAGAKAGIEMLLAQRASIVGAKNSFPAVKATLRVDEPESVRLH
PNATPECALVEGFPVGGFEDPGILRQMRITVRVGEVSTYASVENADLAYEALN
IDRAQMTLISEYPLHDIG"
BASE COUNT      1386 a 1807 c 1681 g 1460 t
ORIGIN
Alignment Scores:
Pred. No.:      5.68e-118      Length:      6334
Score:          2005.00      Matches:      379
Percent Similarity: 100.00%      Conservative: 0
Best Local Similarity: 100.00%      Mismatches: 0
Query Match:    100.00%      Indels:      0
DB:             1      Gaps:      0
US-10-007-527a-2 (1-379) x AY178757 (1-6334)
QY      1 MetThSerValSerAlaGluHisLeuSerGlyValAspArgProProValLeuValSer 20
Db      3051 ATGACCGAGCTAAGTGTGTAACACCTTTCCGCGAAGACCGGCTCCCTCGTGTGTG 2992
QY      21 SerAspIysArgGlyLeuArgHisGluLeuArgProIysLeuGlnIleThrThrSer 40
Db      2991 TCCGATTAAGCGCGGCAATCCGCGACGAACTGCAACCAATCAACACGCTCA 2932
QY      41 GluThrPheAsnIaCysGlyArgProIleSerGlyValAsnGlyValThrIleValaAn 60
Db      2931 GAAACATTTAACCCCTGTGGCCGCGCATTTCTGCGGTGAACCGTGTGACCAATGTGCAAC 2872
QY      61 GlyProIysGlySerGlyPheGlyGlyLeuArgSerGlyIysGlyTyrIleCysPro 80
Db      2871 GGTCCGAAGGTTCTGATTCGAGAGGCTTCGTTCTCGCGAAGAGGCTGATTCGCCCC 2812
QY      81 CysCysAlaGlyIysValGlyAlaHisArgAlaAspGlnIleSerGlnValValaHis 100
Db      2811 TGTGCTGCGGGAAGAGCGGTGCAATCGTGACAGAGAAATTTCTCAAGTGTGTGCTCAT 2752
QY      101 GluLeuGlyThrGlySerValAlaMetValThrMetThrSerArgHisThrAlaGlyGln 120
Db      2751 GAACTCGGAGCTGATCTGTGCGATGTGTACATACATGCGCATACAGCTGAGTCAAG 2692
QY      121 ArgLeuHisAspLeuTyrThrGlyLeuSerAlaAlaTyrIysAlaAlaThrAsnGlyArg 140
Db      2691 CGGCTCCACGACCATGACCTGAGACTTTCCGCGACCTCGAAGAACTGCGACCAACGAGTCT 2632
QY      141 ArgThrArgThrGluArgGluMetTyrGlyCysAsnGlyTyrThrAlaArgAlaGluIle 160
Db      2631 CGTGGCGAGCGGAGACCTGAAATGTACGCTGCGACGAGATACCTGGCGCTGTGTAATTC 2572
QY      161 ThrHisGlyIysAsnGlyTyrHisValHisValHisAlaLeuLeuMetPheSerGlyAsp 180

```

```

Db      2571 ACTACCGAAGAAAACCGCTGGCACGTCCTCACGCTTCAAGCTACTCATGTTCAGTGTGAC 2512
QY      181 ValSerGluAsnIleLeuGlnSerPheSerAspAlaMetPheAspArgTyrThrSerIys 200
Db      2511 GTGAGTAGAAGAACCTCCGAAATCCTTCTCGAGAGCATGTTCATCGGTGACCTTCCAAA 2452
QY      201 LeuValSerLeuGlyPheAlaAlaPolLeuArgAsnSerGlyGlyLeuAspValaGlyIys 220
Db      2451 CTGCTATCTCTCGGATTTGCTGTGCGACCACTACGTAATTCGGGTGTCTCCATGTACAAAAG 2392
QY      221 IleGlyGlyIuAlaAspGlnValLeuAlaAlaTyrLeuThrIysIleAlaSerGlyVal 240
Db      2391 ATCGCGCGTAGAGCTGATTCAGATTTCTCGCTCGCATCTCAAGAAATTCATCTGCGCTT 2332
QY      241 GlyMetGluValGlySerGlyIysArgGlyIysSerGlyArgHisGlyAsnArgAlaProTyr 260
Db      2331 GGTATGAGAGGTGTGTATGTGCGACGAGAAAAGGTGTGCAATGACATGCAACCTGCAACCTCG 2272
QY      261 GluIleAlaValAspAlaValaGlyIysAspProGlnAlaLeuGlnLeuTyrPArgGluPro 280
Db      2271 GAATTCCTCTGTGATGACAGTGGCGGGGATCCACACGCTTGGAACCTGTGGCAGAAATTT 2212
QY      281 GluPheGlySerMetGlyArgArgAlaAlaIleAlaTyrSerArgGlyLeuArgAlaArgAla 300
Db      2211 GAGTTGTGTTTGATGGGACGTCGCGCAATGCGCGGTGTCCTCGATTTGCGTCCGAGCT 2152
QY      301 GlyLeuGlyAlaGluLeuThrAspAlaGlnIleValGluGlnIleGlnIleGlnIleGlnIle 320
Db      2151 GGTCTTGCGGAGCACTAACAGATGCTCAATGCTGTGAGAGAGAGAAATGTGCCCGGTC 2092
QY      321 MetValAlaIleIleProAlaArgSerTyrMetMetIleArgThrCysAlaProTyrVal 340
Db      2091 ATGTTTCGATCATTCATCCGGCGGATCGTGTGATGATGATTCGACTGTGTGCTTACGTC 2032
QY      341 PheGlyGlyIleLeuGlyLeuValGlyAlaGlyAlaThrTyrProIysAsnLeuArgAspHis 360
Db      2031 TTCGGCAGATCCCTCGGACCTGTGCAAGCTGGCGCGCATTTGGGAAAATCTTCGTGTATCAC 1972
QY      361 LeuHisTyrArgLeuProAlaAlaAspValaArgProProIleSerValaArgIys 379
Db      1971 TTGCATTTATCATTTGCCCGCAGCGAGATGTGCGGCCCCCGATATATGTGTTCCGACG 1915
RESULT 3
AX548646/c 6334 bp DNA linear PAT 27-NOV-2002
LOCUS AX548646
DEFINITION Sequence 5 from Patent WO2005709.
ACCESSION AX548646
VERSION AX548646.1 GI:25813614
KEYWORDS
SOURCE
ORGANISM Rhodococcus erythropolis
Rhodococcus erythropolis
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Corynebacterineae; Nocardiaceae; Rhodococcus.
REFERENCE 1. Brumucci, M.G., Cheng, Q., Kostichka, K.N. and Tomb, J.F.
AUTHORS Rhodococcus cloning and expression vectors
TITLE Patent: WO 02055709-A 5 18-JUL-2002,
JOURNAL E.I. DU PONT DE NEMOURS AND COMPANY (US)
FEATURES
source 1..6334
/organism="Rhodococcus erythropolis"
/mol_type="genomic DNA"
/strain="AN12"
/db_xref="taxon:1833"
BASE COUNT 1386 a 1807 c 1681 g 1460 t
ORIGIN
Alignment Scores:
Pred. No.:      5.68e-118      Length:      6334
Score:          2005.00      Matches:      379
Percent Similarity: 100.00%      Conservative: 0
Best Local Similarity: 100.00%      Mismatches: 0

```

Query Match: 100.00% Indels: 0
 DB: 6 Gaps: 0
 US-10-007-527a-2 (1-379) x AX548646 (1-6334)

QY 1 MetThrSerValSerAlaGluHisLeuSerGlyLeuAspArgProValLeuValSer 20
 DB 3051 ATGACGAGCGTAAGTGTGAACACCTTCCGGCAAGAGCGGCTCCCGCTCGGTG 2992

QY 21 SerAspLysArgGlyIleArgHisGluLeuArgProLysLeuGlnIleThrThrSer 40
 DB 2991 TCCGATTAAGCGCGCACTCCGACGAACTGCGACCAACTTCAACAAATCCACGCTA 2932

QY 41 GluThrPheAsnAlaCysGlyArgProIleSerGlyValAsnGlyValThrIleValAsn 60
 DB 2931 GAACATTTAAACGCTGTGCGCCGCGCAATTTCTGGCGTGAACGCTGTGACCTGCAAC 2872

QY 61 GlyProLysGlySerGlyPheGlyGlyLeuArgSerCysGlyLysGlyTyrIleCysPro 80
 DB 2871 GGTCGGAAGGTTCTGGATTCCGAGGCGCTTCCTTCGCGAAGGCGTGGATCTGCCCC 2812

QY 81 CysCysAlaGlyLysValGlyValHisArgAlaAspGluIleSerGlnValAlaHis 100
 DB 2811 TGCTGGCGGAAAGTCGGTGACATGTCGACGAAATTTCTCAAGTTGTGCTCAT 2752

QY 101 GlnLeuGlyThrGlySerValAlaMetValThrMetThrMetArgHisThrAlaGlyGln 120
 DB 2751 CAACTCGGAGCTGGATCTGTTCGATGTGATGACATGACATGCGCATACAGCTGGTCA 2692

QY 121 ArgLeuHisAspLeuTyrThrGlyLeuSerAlaAlaTyrIleValAlaThrAsnGlyArg 140
 DB 2691 CGGCTCCACGACCTATGAGCTGACCTTTCCGACGCTCGAAGAGCTCGACCAACGCTGCT 2632

QY 141 ArgTyrArgThrGluArgGluMetTyrGlyCysAspGlyTyrValArgAlaValGluIle 160
 DB 2631 CGTTGGCGTACGAGCACTGAAATGTACGCGCTGCGACGAGAACCTGCGCTGTGAATC 2572

QY 161 ThrHisGlyLysAsnGlyTyrPheHisValHisValHisAlaLeuLeuMetPheSerGlyAsp 180
 DB 2571 ACTCAGCAAGAAAAAGCGCTGCGACGCTCAGCTCAGCTCAGCTCAGTTCATGCTGATGAC 2512

QY 181 ValSerGluAsnIleLeuGluSerPheSerAspAlaMetPheAspArgTyrThrSerLys 200
 DB 2511 GTGATGAGAACATCTCTCGAATCTTCTCGATGCGATGTTCCATCGGTGAGCTTCCAA 2452

QY 201 LeuValSerLeuGlyPheAlaAlaProLeuArgAsnSerGlyGlyLeuAspValArgLys 220
 DB 2451 CTGCTATCTCGGATTTGCTGCGCCACCTACGTAAATTCGGGTGCTCGATGATGACAAAG 2392

QY 221 IleGlyGlyGluAlaAspGlnValLeuAlaAlaTyrLeuThrLysIleAlaSerGlyVal 240
 DB 2391 ATCGCGGTGAAGCTGATCAAGTCTCGCTGCGATCTGACGAAATTTGATCTGGCGAT 2332

QY 241 GlyMetGluValGlySerGlyAspGlyLysSerGlyValArgHisGlyAsnArgAlaProTyr 260
 DB 2331 GGATGAGAGGTGATGATGAGCGGCAAGAAAGTGTGATGACATGGCAACCTGACACCTGG 2272

QY 261 GluIleAlaValAspAlaValGlyGlyAspProGlnAlaLeuGluLeuTyrPdaGlyLuphe 280
 DB 2271 GAAATCGCTGTGATGAGTGGGCGGGGATCCACAAAGCTTGGAAAGCTGGCAAGAAATT 2212

QY 281 GluIlePheGlySerMetGlyArgArgAlaIleAlaThrSerArgGlyLeuArgAlaArgAla 300
 DB 2211 GAGTTGGTTCGATGGGACGTCGGGCAATGCGGTGCTCCGTTGATTTGCTGCGCCAGCT 2152

QY 301 GlyLeuGlyAlaGluLeuThrAspAlaGlnIleValGluGlnGluGlnSerAlaProVal 320
 DB 2151 GGTCTTGGGCGCAAGATCAACAGATGCTCAATGCTTGAACAGAAAGATTCGCCCGGCTC 2092

QY 321 MetValAlaIleIleProAlaArgSerTyrMetMetIleArgThrCysAlaProTyrVal 340
 DB 2091 ATGATTCGATCATTCGGCGCGGATGATGATGATGATGATGATGATGATGATGATGATG 2032

QY 341 PheGlyGluIleLeuGlyLeuValGluAlaGlyAlaThrTyrGluAsnLeuArgAspHis 360
 DB 2031 TTGGCGAGATCTCTCGACCTGCTCGAAGCTGGGCGCACTTGGGAAATCTTCTGATCAC 1972

QY 361 LeuHisTyrArgLeuProAlaAlaAspValArgProIleIleSerValArgLys 379
 DB 1971 TTGCATTATGATTTGCCGCGAGATGTCGCGCCCGCATATATATGCTTCCGCAAG 1915

RESULT 4

AX548648/c

LOCUS AX548648 9652 bp DNA linear PAT 27-NOV-2002
 DEFINITION Sequence 7 from Patent WO2055709.
 ACCESSION AX548648
 VERSION AX548648.1 GI:25813616

KEYWORDS
 SOURCE
 ORGANISM
 Shuttle vector pRHR17
 Shuttle vector pRHR17
 artificial sequences; vectors.

REFERENCE
 1 Brumucci, M.G., Cheng, Q., Kostichka, K.N. and Tomb, J.F.
 "RhoDccoccus cloning and expression vectors
 Patent: WO 02055709-A 7 18-JUL-2002;
 E.I. DU PONT DE NEMOURS AND COMPANY (US)

FEATURES
 source
 1..9652
 /organism="Shuttle vector pRHR17"
 /mol_type="genomic DNA"
 /db_xref="taxon:214938"

BASE COUNT 2097 a 2729 c 2559 g 2267 t
 ORIGIN

Alignment Scores:
 Pred. No.: 9.26e-118 Length: 9652
 Score: 2005.00 Matches: 379
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 6 Gaps: 0

US-10-007-527a-2 (1-379) x AX548646 (1-9652)

QY 1 MetThrSerValSerAlaGluHisLeuSerGlyLysAspArgProValLeuValSer 20
 DB 5052 ATGACGAGCGTAAGTGTGAACACCTTCCGCAAGAGCGGCTCCCGCTCGGTG 4993

QY 21 SerAspLysArgGlyIleArgHisGluLeuArgProLysLeuGlnIleThrThrSer 40
 DB 4992 TCCGATTAAGCGCGCATCCGCGCAAGACTGCGACCAACTTCAACAAATCACCACGTCA 4933

QY 41 GluThrPheAsnAlaCysGlyArgProIleSerGlyValAsnGlyValThrIleValAsn 60
 DB 4932 GAACATTTAAACGCTGTGGCGCGCGCAATTTCTGACGTGAACGCTGACCATTTGTCAC 4873

QY 61 GlyProLysGlySerGlyPheGlyGlyLeuArgSerCysGlyLysGlyTyrIleCysPro 80
 DB 4872 GGTCGGAAGGTTCTGATTTGAGAGGCTTGTTCGCGGAAAGGCTGGATCTGCCCC 4813

QY 81 CysCysAlaGlyLysValGlyValHisArgAlaAspGluIleSerGlnValAlaHis 100
 DB 4812 TGCTGTGGGAAAGTCGGTGACATGTCGACACCAAAATTTCTCAAGTTGTTCAT 4753

QY 101 GlnLeuGlyThrGlySerValAlaMetValThrMetThrMetArgHisThrAlaGlyGln 120
 DB 4752 CAACCTCGGAGCTGGATCTGTTGCGATGTGACATGACATGACATGACATGACATGAC 4693

QY 121 ArgLeuHisAspLeuTyrThrGlyLeuSerAlaAlaTyrIleValAlaThrAsnGlyArg 140
 DB 4692 CGGCTCCAGACCTATGACTGACTTTCGCGACCTCGGAAAGCTGGACCAACGATCGT 4633

QY 141 ArgTyrArgThrGluArgGluMetTyrGlyCysAspGlyTyrValArgAlaValGluIle 160
 DB 4632 GGTTCGCTACGAAACGTAATATACGCTGCGACGAGATGATGATGATGATGATGATGATG 4573

Db 5621 TTGGCGGAGATCTCGGACTCGTCGGAAGCTGGGCGGACTTGGGAAAATCTTCGTGATCAAC 5562
Qy 361 LeuHISTYArgLeuProAlaAlaAspValArgProProIleIleSerValArgLys 379
Db 5561 TTGCATTATGATTGCCCCGAGCGGATGTGGGCCCCCGCATATATATGATTGCAAG 5505

RESULT 6
AY180162/c 11241 bp DNA circular SYN 01-MAY-2003
LOCUS Shuttle vector pRHBRI7, complete sequence.
DEFINITION AY180162
ACCESSION AY180162
VERSION AY180162.1 GI:30313707
KEYWORDS
SOURCE Shuttle vector pRHBRI7
ORGANISM Shuttle vector pRHBRI7
REFERENCE 1 (bases 1 to 11241)
AUTHORS Kostichka, K., Tao, L., Brammucci, M., Tomb, J.-F., Nagaraajan, V. and Cheng, Q.
TITLE A small cryptic plasmid from Rhodococcus erythropolis: characterization and utility for gene expression
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 11241)
AUTHORS Kostichka, K., Tao, L., Brammucci, M., Tomb, J.-F., Nagaraajan, V. and Cheng, Q.
TITLE Direct Submission
REFERENCE Submitted (14-NOV-2002) CR@P, E. I. Dupont de Nemours Inc., Experimental Station, Wilmington, DE 19880-0328, USA
FEATURES
LOCATION/Qualifiers
1..11241
/organism="Shuttle vector pRHBRI7"
/mol_type="genomic DNA"
/db_xref="taxon:214938"
/focus
/note="Rhodococcus - Escherichia coli shuttle vector: Ssp I linearize PAN12 cloned into the PvuII site of pBR328"
join(9925..11241,1..3590)
/organism="Cloning vector pBR328"
/mol_type="genomic DNA"
/db_xref="taxon:221946"
3591..9924
/organism="Rhodococcus erythropolis"
/mol_type="genomic DNA"
/db_xref="taxon:1833"
/plasmid="PAN12"
/note="cryptic plasmid"

BASE COUNT 2517 a 3109 c 2945 g 2670 t
ORIGIN

Alignment Scores:
Pred. NO.: 1.1e-117 Length: 11241
Score: 2005.00 Matches: 379
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 12 Gaps: 0

US-10-007-527a-2 (1-379) x AY180162 (1-11241)

Qy 1 MetHISerValSerAlaGluHisLeuSerGlyLysAspArgProProValIleuValSer 20
Db 6641 ATGACCAAGCGTAATGTGTGAACACCTTCCGGCAAGACCGGCTCCCGTCCGTGTG 6582
Qy 21 SerAspLysArgGlyIleArgHisGluLeuArgProLysLeuGlnGlnIleThrThrSer 40
Db 6581 TCCGATTAAGCGGCAATCCGCAAGAACTCGAAGCCCAACTTCACAAATCACACAGCTCA 6522
Qy 41 GluThrPheAsnAlaCysGlyArgProIleSerGlyValAsnGlyValThrIleValAsn 60
Db 6521 GAATCATTTAAACGCTGTGGCCGCGCATTTCTGGCTGTAACGCTGTGACCATTTGCAAC 6462
Qy 61 GlyProLysGlySerGlyPheGlyLysLeuArgSerGlyLysGlyTrpIleCysPro 80

Db 6461 GGTCCGAAAGATTCTGGAATTCGAGGCTTCGTCTCGGAAAAGGCTGATTCGCCCC 6402
Qy 81 CysCysAlaGlyLysValGlyAlaHisArgAlaAspGluIleSerGlnValAlaHis 100
Db 6401 TGCCTGCGGAAAAGCGGTGACATCGTCGAGAGAAATTTCTTAATTTGTGCTCAT 6342

Qy 101 GlnLeuGlyThrGlySerValAlaMetValThrMetThrMetArgHisThrAlaGlyGln 120
Db 6341 CAATCGGAGCTGATATGTTGCGATGATGATGATGATGATGATGATGATGATGATGATG 6282
Qy 121 ArgLeuHisAspLeuTrpThrGlyLysSerAlaAlaTrpLysAlaAlaThrAsnGlyArg 140
Db 6281 CGGCTCCACCACTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 6222

Qy 141 ArgTrpArgThrGluArgGluMetTrpGlyCysAspGlyTrpValArgAlaGluIle 160
Db 6221 CGTGGCGTACGGAACCTGAAGATGATGATGATGATGATGATGATGATGATGATGATG 6162
Qy 161 ThrHisGlyLysAsnGlyTrpHisValHisValHisAlaLeuLeuMetPheSerGlyAsp 180
Db 6161 ACTCAGGAAAACCGGCTGCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCA 6102

Qy 181 ValSerGluAsnIleLeuGluSerPheSerAspAlaMetPheAspArgTrpThrSerLys 200
Db 6101 GTGATGAGAACATCTCGAATCTCTCGAATGCGATGTTGATGATGATGATGATGATGATG 6042
Qy 201 LeuValSerLeuGlyPheAlaAlaProLeuArgAsnSerGlyLysAspValArgLys 220
Db 6041 CTCGTATCTCTGGGATTTGCTGCGCCCACTACGTAATTCGATGATGATGATGATGATG 5982

Qy 221 IleGlyGlyGluAlaAspGlnValLeuAlaAlaTrpLeuThrLysIleAlaSerGlyVal 240
Db 5981 ATCCGCGGTGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 5922
Qy 241 GlnMetGluValGlySerGlyLysAspGlyLysSerGlyArgHisGlyAsnArgAlaProTrp 260
Db 5921 GGTATGAGGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 5862

Qy 261 GluIleAlaValAlaAspAlaValGlyLysProGlnAlaLeuGluLeuTrpArgGluPhe 280
Db 5861 GAATTCCTGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 5802
Qy 281 GluPheGlySerMetGlyArgArgAlaIleAlaTrpSerArgGlyLeuAlaAlaGlyAla 300
Db 5801 GAGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 5742

Qy 301 GlyLeuGlyAlaGluLeuThrAspAlaGlnIleValGluGlnGlnGlnSerAlaProVal 320
Db 5741 GGTCTGGGCGGAGAACTTAACAGATGCTCAATGCTTGAAGAGAGATGCTGCCCGCTC 5682
Qy 321 MetValAlaIleIleProAlaArgSerTrpMetMetIleArgThrCysAlaProTrpVal 340
Db 5681 ATGATGAGAACTTCCGCGGCGATCTGATGATGATGATGATGATGATGATGATGATGATG 5622

Qy 341 PheGlyGluIleLeuGlyLysValGluAlaGlyAlaThrTrpGluLeuLeuArgAspHis 360
Db 5621 TTCCGCGAGATCTCGAATCTGTAAGCTGCGCGCACTTGGGAAAATCTTCGTGATCAC 5562
Qy 361 LeuHISTYArgLeuProAlaAlaAspValArgProProIleIleSerValArgLys 379
Db 5561 TTGCATTATGATTGCCCCGAGCGGATGTGGGCCCCCGCATATATATGATTGCAAG 5505

RESULT 7
AP083788 2439 bp DNA circular BCT 15-JUN-1998
LOCUS Actinomyces pyogenes plasmid pAP1, complete plasmid sequence.
DEFINITION U83788
ACCESSION U83788
VERSION U83788.1 GI:1805288
KEYWORDS
SOURCE Arcanobacterium pyogenes
ORGANISM Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;

REFERENCE 1 Actinomycetaceae; Actinomycetaceae; Arcanobacterium.
 1 (bases 1 to 2439)
 AUTHORS Billington, S.J., Jost, B.H. and Songer, J.G.
 TITLE The Arcanobacterium (Actinomycet) pyogenes plasmid papi is a member
 of the pU101/pU1 family of rolling circle replication plasmids
 JOURNAL U. Bacteriol. 180 (12), 3233-3236 (1998)
 MEDLINE 98292760
 PUBMED 9620977
 REFERENCE 2 (bases 1 to 2439)
 AUTHORS Billington, S.J., Jost, B.H. and Songer, J.G.
 TITLE Direct Submission
 JOURNAL Submitted (06-JAN-1997) Veterinary Science, University of Arizona,
 1117 East Lowell Street, Tucson, AZ 85721, USA
 FEATURES
 SOURCE 1. .2439
 /organism="Arcanobacterium pyogenes"
 /mol_type="genomic DNA"
 /strain="BB1"
 /db_xref="taxon:1661"
 /plasmid="papi"
 1. .110
 /note="single stranded origin; putative"
 39. .44
 /note="size of single strand initiation; putative"
 221. .1600
 /gene="rep"
 221. .1600
 /gene="rep"
 /function="replication protein"
 /codon_start=1
 /transl_table=11
 /product="Rep"
 /protein_id="AAC46399.1"
 /db_xref="GI:1805289"
 /translation="MNRISRTALSLPARIQKIPPAAGSRSLKSEFGMTATWSARG
 ASSDSERDRSOPSNRRRSGSATHTLGTATLTPVNSKKTAKRSERYELRDG
 LAESTIESVRKGRVPAPVAPVSLRAKSDKAGVGGITGCGWASCPICSAKIAAR
 KTDIQVVDHVAKHGVTSMVTLTORHKGQGLKHLMDALSTAMNVTGGRNTEKE
 QFGLVYVRANETIHKGGWVHSHVLIISKDPLSTFYORQGRRLPYPELTM
 SDFIARWRAEGLAKGVDFLSDGGIDMTAKADARAIGNVYKQTSIDASVETL
 GPFKAKNGRTPFOILADILSLGDVDLTLKREYKASFGRRALTWKGLRDMAIG
 VEOGDEETIASEBIDEALFLTHDAMQVRRFGAELLDTVESGGRALVRYMLDFREI
 DMSLPKITE"

RBS 1673. .1676
 1685. .2023
 /note="ORF112"
 /codon_start=1
 /transl_table=11
 /product="unknown"
 /protein_id="AAC46400.1"
 /db_xref="GI:1805290"
 /translation="MFTIGKRTKFOPIADCKSPFDNGKRTGSESEKMTQAIT
 EDGLVETLSIPFOVOLLIDAVKSRLEEFKVNLAIVLRNGGISTYVLTGEAVHV
 PAPVVEVSE"
 2010. .2013
 2020. .2409
 /note="ORF129"
 /codon_start=1
 /transl_table=11
 /product="unknown"
 /protein_id="AAC46401.1"
 /db_xref="GI:1805291"
 /translation="MMSBPYGISNATLOLVSPRPAKRAEATYMSRIKKELEGKHS
 TPTALASCEKTVSVEPRIVDVVHLHEVGECCACAGNSAVASLSRAFTIPARFQS
 LENGRRNPQQLITQDVAGMEKTLGTR"
 2378. .2406
 /note="double stranded origin; putative"
 BASE COUNT 603 a 657 c 656 g 523 t
 ORIGIN

Alignment Scores:
 Pred. No.: 7,77e-30 Length: 2439
 Score: 609.50 Matches: 148

Percent Similarity: 51.47% Conservative: 62
 Best Local Similarity: 36.27% Mismatches: 133
 Query Match: 30.40% Indels: 65
 DB: 1 Gaps: 15
 US-10-007-527a-2 (1-379) x AP083788 (1-2439)
 QY 5 SerAlaGluHisLeuSerGlyIlyAspArgProProValLeuValSerSerAsp----- 22
 DB 419 TCAGGACCCATCCCTTGGCAATACGATTTACTTTCAGATCAAGAGAGAGAG 478
 QY 23 -----LysArgGlyIleArgHisGluLeuArgProLysGluGlnIle 37
 DB 479 AAAACAGCTAATCTCCGCGCTTCTGAGAGATACCACTCGAGACGAGTTCGGAATC 538
 QY 38 ThrThrSerGluThrPheAsnAlaCysGlyArg---ProIleSerGlyValAsnGlyVal 56
 DB 539 TCAGCATTTGATGCTCGTCCGAAAGTGGCCGCGTCCGCGGACCTCTGCTCGTTG 598
 QY 57 ThrIleValAsnGlyProLysGlySerGlyPheGlyIleLeuArgSerCysGlyLysGly 76
 DB 599 CGAGCAAAATCTGACGGTAAAGCGCGGATATGCTGTTGACACACTGTGGAAGGCTC 658
 QY 77 TrpIleCysProCysCysAlaGlyLysValGlyValHisArgAlaAspGluIleSerGln 96
 DB 659 TGGGCTCCCACTGCTGTAGCCGAAATCCGCGCTGCGGAAAACCGACTTCACACG 718
 QY 97 ValValAlaHisGluLeuGlyThrGly---SerValAlaMetValThrMetThrMetArg 115
 DB 719 GTGCTTGACCAAGCGGTAAACAGGATGACCGTTCATGCTTACGCTCACCAAGCT 778
 QY 116 HisThrAlaGlyGlnArgLeuHisAspLeuThrGlyLeuSerAlaAlaTrpLysAla 135
 DB 779 CACCAAGAGCAAGAGGCTAAACACCTCTGGACGCGCTTGTGACGCGCTGGAATCGC 838
 QY 136 AlaThrAsnGlyArgArgTrpArgThrGluArgGluMetTrpGlyCysAspGlyValVal 155
 DB 839 GTTACCTCTGCTGCTGCTGATTTGATTCAGAGCAATTTGGTTAGTTCGTTAGTT 898
 QY 156 ArgAlaValGluIleThrHisGlyLysAsnGlyTrpHisValHisValHisAlaLeuLeu 175
 DB 899 CGAGCAATGAATTAATCATGAGAAAGCAGCGCTGAGTGCATTCCTCATGTTCTGATT 958
 QY 176 MetPheSerGlyAspValSerGluAsnIleLeuGluSerPhe----- 189
 DB 959 ATTTCGAGAAAGAC-----CGCTGACCAAGCACTTGTCTATCAACGAAACAA 1009
 QY 190 -----SerAspAlaMetPheAsp 195
 DB 1010 GAGCGCGCGCGCTTCCCTACCCCGCAGAGTTATATGTCATCGAATTCATTCGAGAA 1069
 QY 196 ArgTrpThrSerLysLeuValSerLeuGlyPheAlaAlaProLeuArgAsnSerGlyGly 215
 DB 1070 CGGTGGAAAGCTGGCGTTGCCAAGACCGCGCTTGAATTT---CTCCGGAATTCGAGGCG 1126
 QY 216 Leu-----AspValArgLysIleGlyGlyGluAlaAspGlnValLeu 229
 DB 1127 TTGACGTGACCGTTGCCAAGAGCGCGGACCACTTCGCAAC----- 1168
 QY 230 AlaAlaIlyLeuThrLysIle-----AlaSerGlyValGlyMetGluValGlySer 246
 DB 1169 -----ThrGTCAGCAAGATGCAAGCAGTCCACAGAGGATTAAGTCCGAGAGTCA 1222
 QY 247 GlyAspGlyLysSerGlyArgHisGlyLysAsnArgAlaProTrpGluIleAlaValAspAla 266
 DB 1223 GCGCGCTTCAAAAAGCCGGAACCGGAAACGAGACCGCTTCCAGATACCTCGCGGATATC 1282
 QY 267 ValGly---GlyAspProGlnAlaLeuGluIleuThrArgGluPheGluPheGlySerMet 285
 DB 1283 CTTGCGTCCGCGAGTGTGAGAGCACTCAAGCTCTGGAAGAAGATATGAGAAAGCTTCGTT 1342
 QY 286 GlyArgArgAlaIleAlaTrpSerArgGlyLeuArgAlaArgAlaGlyLeuGlyAlaGlu 305

Db 1343 GAGCGCGTGCCTTACATGTCGAAAGGGCTCAGAGATTGGCAATCTCGCGCTTGAA 1402
QY 306 LeuThrAspAlaGlnIleValGluGlnGlnIserAlaProValMetValAlaIleIle 325
Db 1403 CAGTCCGACGACGAGATGGCC--TCTGAGGAATCGGAGACGACGATAGCGCTATT 1459
QY 326 ProAlaArgSerThrMetIleArgThrCysAlaProIleValPheGly-----Glu 343
Db 1460 ACCGATGACGCTTGGCGCTCAGTGGCAGCT-----TTTGGAGCGCGCTGAA 1504
QY 344 IleLeuGlyLeuValGluAlaGlyAla-----ThrTrpGluAsnLeuAsp 358
Db 1505 CTCTCGATGTCGACCGAATCCGAGAGCTGCGCGCGCTTACCGCTGTTGATTTTAA 1564
QY 359 AspHisLeuHisIleValArgLeuPro 366
Db 1565 GAA--ATTGATTGTCATTGCT 1585

RESULT 8
LOCUS AY150274
DEFINITION Propionibacterium granulosum cryptic plasmid pPG01, complete
ACCESSION AY150274
VERSION AY150274.1 GI:27465054
KEYWORDS
SOURCE Propionibacterium granulosum
ORGANISM Propionibacterium granulosum
Bacteria; Actinobacteria; Propionibacteriaceae; Propionibacterium.
1 (bases 1 to 3540)
REFERENCE 1 Farrar, M.D. and Holland, K.T.
TITLE Isolation and characterisation of a cryptic plasmid from the human
JOURNAL skin commensal Propionibacterium granulosum
AUTHORS Farrar, M.D. and Holland, K.T.
REVIEWER Unpublished
AUTHORS 2 (bases 1 to 3540)
TITLE Direct Submission
JOURNAL Submitted (13-SEP-2002) Skin Research Centre, Division of
Microbiology, University of Leeds, Leeds LS2 9JT, UK
FEATURES
source
1. 3540
/organism="Propionibacterium granulosum"
/mol_type="genomic DNA"
/strain="PF283"
/db_xref="taxon:33011"
/plasmid="cryptic plasmid pPG01"
74..280
/codon_start=1
/transl_table=11
/product="orf3"
/protein_id="AA078121.1"
/db_xref="GI:27465054"
/translation="MGQGYGILNRGPFYNGSHNEIIGYPMNRVITWPPALLSMPL
SSGFSRGHYVERPRAVEFHCADHD"
449..1792
/gene="tra"
449..1792
/gene="tra"
/note="plasmid transfer protein"
/codon_start=1
/transl_table=11
/product="tra"
/protein_id="AA078122.1"
/db_xref="GI:27465056"
/translation="NAGERTVLTGVYMWLRILIPVLLIGVAFACMCFSHLYVYVIGFA
VAALAEAPSPVWQVPMOYPARVYVWWSIAKAFPISSYGGDRITYRPGIOMI
RSDRHVHTLVLPVPAGLASAVILEKAAETORQLRGKSMRTCTYKPAHGLDIPR
DTRAGDELTPASTISWNVFVYKPDGSSVLDLSPSHIVSGKTRSGKSFYGL
DQMRHLPTVAAGDPTGILLFNLGGMGGDALRSKITNDADAAVVOYLISMTIEMD
RRIYLLNCHRDKMSRNPESDPGRLLIIVILEYVGLIERQNPDSARAGASRFA
SKAAGVGRIVAEKAGVAVLLVTRPAKTIIGPLRAQLTRVTFADSDPMHSMH
PLSSQYVQKQNSMASGVETIADVIFILTRFRSYRAHELTDLRPAASVQIDLLIG"

gene 1979..3280
/gene="rep"
CDS 1979..3280
/gene="rep"
/note="plasmid replication protein"
/codon_start=1
/transl_table=11
/product="Rep"
/protein_id="AA078123.1"
/db_xref="GI:27465057"
/translation="MLNTVTSKQPHRLCSNHRPCARSERQRAAGDRIGTTBAT
PFVSDPKRPNRGRNRYEMEDGLRNQVMELEVRKCGAVPSQRLALMAGHGAG
YAGLATGSGWACPVCAKISARHDELARVQVAVGLGFKVSMLTLTQRHAGDPLA
ELMALGSGWNAVTSGRMOERCAQLGQWAKAVATVTHSGHVMHYVIVISKQDP
TSVDTKIRHRKQGRRTPEYEPVQVRPDLIERSRGLRKRGVNFIAGSGGLDMQTA
DSGDEBALGRYVAKNNSVDGLANRATGAPFKARGNRTPQIILEDPLTSERTDR
LMRTYVSAHGRKALTWSKGLDMAGMSBMSDEYVADQCEAVALLFDHAKRQIR
TAGAFLDELDELHSGSEVYALWLRKRRIHYELPLVWSTST"

BASE COUNT 693 a 967 c 1066 g 814 t
ORIGIN

Alignment Scores:
Pred. No.: 4,46e-27 Length: 3540
Score: 559.00 Matches: 141
Percent Similarity: 51.78% Conservative: 63
Best Local Similarity: 35.79% Mismatches: 152
Query Match: 28.38% Indels: 39
DB: 1 Gaps: 14

US-10-007-527a-2 (1-379) x AY150274 (1-3540)

QY 4 ValSerAlaGlnHisLeuSerGlyLysAspArgProProValLeuValSerAspLys 23
Db 2094 GACACAGCGAGCAACTTTCCGGATACAGATTCCCGCTGAGA--AAGCGAATGAGGCG 2152
QY 24 ArgGlyIleArgHisGlnLeuArgProLysLeuGln-----GlnIleThrSerGlu 41
Db 2153 CGTCGACATCGCTAAGATGAGATGAGGATGATGATACCAAAATCCGACAGTGCAGCG 2212
QY 42 ThrPheAsnAlaCysGlyArg--ProIleSerGlyValAsnGlyValIleThrIleValAsn 60
Db 2213 CGCGTTCGCAAGTGGCGGCGGACGTGCGGTTG-----CAACGATCGCTTGATGCGG 2266
QY 61 GlyProLysGlySerGlyPheGlyIleLeuArgSerGlyLysGlyTrpIleCysPro 80
Db 2267 GGTTCATGCGGTCGCGGTTATGCGGTTATGGGAGAGTGGAGAGTGGCGCTTGCCT 2326
QY 81 CysCysAlaGlyLysValGlyAlaHisArgAlaAspGluIleSerGlnValAlaHis 100
Db 2327 GTCTGTGCGGCAAAATTTCCGCGCACCGCTCGTATGAGCTGGCCCGTGTTCAGAGTT 2386
QY 101 GlnLeuGlyThrGly--SerValAlaMetValThrMetThrMetArgHisThrAlaGly 119
Db 2387 GCGGTTCGACTCGCTTCAAGTGTGCAATGCTGCAAGCTTACTCAACGTCATCATGCTGCT 2446
QY 120 GlnArgLeuHisAspLeuTrpThrGlyLeuSerAlaAlaTrpValAlaHisArgHis 139
Db 2447 CAGATCTCGCGACGCTGTGGCGCTGCTCCAGCTGGGTTGAAGTGTGTCACGAGTGT 2506
QY 140 ArgArgTrpArgThrGlnArgGlnMetTrpGlyCysAspGlyTrpValArgAlaValGlu 159
Db 2507 CGACGCTGACGAAATTTGGCGCTCAGCTCGCGCGCCAGGAGATGGTCAAGCAGTTGA 2566
QY 160 IleThrHisGlyLysAsnGlyTrpHisValHisValHisAlaLeuLeuMetPheSerGly 179
Db 2567 GTCACCCATGGTCCGACGCGGTGACCTTACGACGTCACGTCGCTGTCTCTAAGCAG 2626
QY 180 Asp-----ValSerGlnAsnIle----- 185
Db 2627 GATCCGACTAGCGTTGACACTAAGATTCGGCATCGCCGCAACAGAGTCGCGCGGAC 2686
QY 186 -----LeuGlnSerPheSerAlaMetPheAspArgTrpThrSerLys 200

Db 2687 CCGATTCACAGAGGATACAGAGGCCGGAAGACTTCATCGTGAACGGTGTGCGAGGT 2746

Qy 201 LeuValSerLeuGly-----PheAlaIaIaProLeuArgAsnSerGlyGlyLeuAspVal 218

Db 2747 TTGGAGAGCGCGCGCTCGACTTCATCGCC-----GTMGTGTGTGCTTCGATTGG 2797

Qy 219 ArgPylIleGlyGlyGlyAlaAspGlnValLeuAlaAlaIleuThrIleAspIleAsp 238

Db 2798 CAGACTCGTATTCGTGAGACGAGAAAGCTCTCGGTGCGAGCTCGCAAGATGAACTCG 2857

Qy 239 -----GlyValGlyMetGlyValGlySerGlyAspGlyIleYsSerGlyArgHisGly 255

Db 2858 TCCGTGATGATCGTGAAGCAAGGACCGTGGCGGCTTCAAGAGAGCTGTAGAGGT 2917

Qy 256 AsnArgAlaProTrpGlnIleAlaValAspAlaValGly---GlyAspProGlnAlaIleu 274

Db 2918 AATGAGACCGCTTCCAGATCTCGAAGATTCCTGTGATACGGGCTCGAGACTGCTG 2977

Qy 275 GluLeuTrpArgGluPheGluPheGlySerMetGlyArgArgAlaIleAlaTrpSerArg 294

Db 2978 AGACTCGGCTACCTATGTTCTGCAAGTCAATGCGCGTGAAGCATTTGACGTGCAAG 3037

Qy 295 GlyLeuArgAlaArgAlaGlyLeuGlyAlaGlyLeuThrAspAlaGlnIleValGluGln 314

Db 3038 GTTTCGCTACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3097

Qy 315 GluGluSerAlaProValMetValAlaIleIleProAlaArgSerTrpMetMetIleArg 334

Db 3098 GACCAAGTCGGGGAGAGG---GTGCGCTTTTGGACGATGACGATGACGCGCGCCCA 3154

Qy 335 ThrCys---AlaProTrpValPheGlyGlyIle---LeuGlyLeuValGlyAlaGlyAla 352

Db 3155 ACTGCGCGTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3214

Qy 353 ThrTrpGluLeuLeuArgAspHisIleuHisIleuArgLeuPro 366

Db 3215 GCTTGCTGGAAGAGCA---AGAATCATTAATGAAATACCT 3253

RESULT 9

PF6662 2051 bp DNA circular BCT 08-JUN-1998

LOCUS Propionibacterium freudenreichii plasmid pLME108 rep gene.

DEFINITION A000662 GI:3212127

ACCESSION A000662.1 GI:3212127

VERSION putative; rep gene.

KEYWORDS Propionibacterium freudenreichii

SOURCE Propionibacterium freudenreichii

ORGANISM Bacteria; Actinobacteria; Actinobacteriales; Actinomycetales; Propionibacteriaceae; Propionibacterium.

REFERENCE 1

AUTHORS Dasen,G.H., Miescher,S., Teuber,M. and Welle,L.

TITLE Molecular analysis of plasmid pLME108, a plasmid isolated from

JOURNAL Unpublished

AUTHORS Dasen,G.H.

TITLE Direct Submission

JOURNAL Submitted (03-JUN-1998) Dasen G.H., Institute of Food Science,

AUTHORS Laboratory of Food Microbiology, ETH Zurich, LFO G24.2,

JOURNAL Schmelzberger, 9, CH-8092 Zurich, SWITZERLAND

FEATURES

SOURCE location/Qualifiers

1..2051

/organism="Propionibacterium freudenreichii"

/mol_type="genomic DNA"

/db_xref="taxon:1744"

/lab_host="DP2"

/plasmid="pLME108"

/complement(223..1434)

/gene="rep"

/complement(1223..1434)

/gene="rep"

/codon_start=1

/transl_table=11

BASE COUNT 427 a 777 c 530 g 317 t

ORIGIN

Alignment Scores:

Pred. No.: 8.86e-23 Length: 2051

Score: 497.00 Matches: 127

Percent Similarity: 50.80% Conservative: 63

Best Local Similarity: 33.96% Mismatches: 144

Query Match: 24.79% Indels: 40

DB: 1 Gaps: 15

US-10-007-527a-2 (1-379) x PF6662 (1-2051)

Qy 19 ValSerSerAspIleValGlyIleArgHisGlyLeuArgProIleGlnGlnIleThr 38

Db 1296 GTGCGCTGTGCAAGACCGCAACACCGGTTTTCGCTTCCGTTATGTTGTTGGCGCTCAACG 1237

Qy 39 ThrSerGlnThrPheAsnAlaCysGlyArg---ProIleSerGlyValAsnGlyValThr 57

Db 1236 TCGTGAAGCGGGTTCGCTTCGCTGCGTGTGTCGCCCTCGCGGCTCGCTCGTGTGT 1177

Qy 58 IleValAsnGlyProIleGlySerGlyPheGlyValGlyLeuArgSerGlyIleValThr 77

Db 1176 GTGCGGTGACGACGACGCGCGCGCGCTTTCGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1117

Qy 78 IleCysProCysCysAlaGlyIleValGlyAlaIleArgAlaAspGlnIleSerGlnVal 97

Db 1116 GCTTCGCGCGGTGTGTAACGCAAGATCATGCTGCGGTGCTTGAATTCGCGCGCGCG 1057

Qy 98 Val-----AlaHisGlnLeuGlyThrGlySerValAlaMetValThrMetThr 113

Db 1056 GTCCAGACCTGAGACTAAGCAC-----GTTGTGTGTGTGTGCTTCATGACTTCACG 1006

Qy 114 MetArgHisIleThrAlaGlyGlnArgLeuHisAspLeuTrpThrGlyLeuSerAlaIleTrp 133

Db 1005 GTTCGCACTCTCGTAAGATTCCTTGAACGCTGTTGGATGATGCTGCTGCGGGTGG 946

Qy 134 LysAlaAlaIleThrAsnGlyArgArgTrpArgThrGlnArgIleMetTrpGlyCysAspGly 153

Db 945 CGGCGAGTCACCTCGGGCAAGGCTGACGCTTCAACGCTCCGCTCCAGGGGTGAGGCG 886

Qy 154 TyrValArgAlaValAlaGlyIleThrHisGlyIleAsnGlyTrpHisValHisValHisAla 173

Db 885 TTCGTGCGTGTGTGAGATTTACGACGCGGCAATGATGATGATGATGATGATGATGATG 826

Qy 174 LeuLeuMetPheSerGlyAspValSerGlnAsnIleLeuGlnSerPheSerAspAlaMet 193

Db 825 CTCGTGCTTCGTGTGCGGACCTTCGCGACGCTGTG-----GGCTGCAACCGGTGCATG 772

Qy 194 PheAspArgTrpThrSerIleValSerLeuValSerLeuGlyPheAlaIleProLeuArgAsn 213

Db 771 TTCGCGCGTTGGGAACGCGCGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCGG 712

Qy 214 GlyGlyLeuAspValArgLysIleGlyGlyAlaAsp---GlnValIleuAlaIleVal 232

Db 711 -----CAAGAGCTTACAGACATGACGCAACCACTGCTGATTCACCTGCGCGCTAC 658

Qy 233 LeuThrIle-----IleAlaSerGlyValGlyMetGlyValGlySerGlyAspGlyLys 250

Db 657 CTGACCAAGCGCACTTCAGGGGAAGATCGGCAAGATTGACGAATCCCAAGAGCAAG 598

QY 251 SerGIYArgHis-----GlyAsnArgAlaProTyrGluIleAlaValAlaAspAlaValGly 268
 Db 597 ACCGCCGCTTCTCGCCTCTCGACTCGATCGACTGAGAGTCTCTGGCGGCGGCGCAAC 538
 QY 269 GlyAspProGlnAlaLeuGluIleuTyrPargGluPheGluPheGlySerMetGIYArgArg 288
 Db 537 GGTTCGGCCGAAGAAGTGTCTCTGTTCGAGTGGAGAGGTTAGCCGGGAGCCAG 478
 QY 289 AlaIleAlaTyrSerArgGlyLeuArgAlaArgAlaGlyLeuGlyValaGluLeuThrAsp 308
 Db 477 CAGATTGGCTGTCTCGGGGTTGAGGAGACGCTCTGGCTTACGTGTAAGAAATCGAC 418
 QY 309 AlaGlnIleValaGluGlnGlu---GluSerAlaProValMetValAlaIleIleProAla 327
 Db 417 GACGTATCGCGCTGAGAGAGTGGCGGTCTGTTCGACACTGCTCCCTTGTATCACCGGCT 358
 QY 328 ArgSerTyrMet-----MetIleArgThrCysAla 337
 Db 357 GACGATGGCGTCTCGCTCGGGGCGACAGCAAGCTTACGAGTGTCTGGAGAGTTC--- 301
 QY 338 ProTyrValPheGlyGluIleLeuGlyLeuValaGluAlaGly---AlaThrTyrGluAsn 356
 Db 300 -----GAATTGGCGCGT-----CAAGCTGGCTTGGCGCGGTGG----- 268
 QY 357 LeuArgAspHis---LeuHisTyrArgLeuProAlaAlaAsp 369
 Db 267 CTGCAGAGCAGCAGATCGAACACGATTGTGTGAGGTGAC 226

RESULT 10
 SLMPG33A 2661 bp DNA linear BCT 09-MAY-1994
 LOCUS S.lavendulae miniplasmid pSLG33 sequence.
 DEFINITION X69872
 VERSION X69872.1 GI:47179
 KEYWORDS miniplasmid.
 SOURCE Streptomyces lavendulae
 ORGANISM Streptomyces lavendulae
 Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 Streptomycinae; Streptomycetaceae; Streptomyces.

REFERENCE
 1 Felsberg,J., Petricek,M. and Richy,P.
 Nucleotide sequence of the mini-plasmid pSLG33 from Streptomyces
 lavendulae-grasseri RIA746
 JOURNAL Nucleic Acids Res. 21 (15), 3582 (1993)
 MEDLINE 93348001
 PUBMED 8346038
 REMARK (sites)
 2 (bases 1 to 2661)
 Felsberg,J.
 TITLE Direct Submission
 SUBMITTED (28-DEC-1992) J. Felsberg, Inst. of Microbiology, Cell
 and Molecular Microbiology Div., Videnska 1083, Prague 4,
 CZECHOSLOVAKIA

FEATURES
 source location/Qualifiers
 1..2661
 /organism="Streptomyces lavendulae"
 /mol_type="genomic DNA"
 /strain="RIA746"
 /db_xref="taxon:1914"
 /dev_stage="mycelia"
 /plasmid="miniplasmid pSLG33"

BASE COUNT 514 a 916 c 818 g 413 t
 ORIGIN

Alignment Scores:
 Pred. No.: 7 456-20 Length: 2661
 Score: 453.00 Matches: 129
 Percent Similarity: 48.33% Conservative: 59
 Best Local Similarity: 33.16% Mismatches: 160
 Query Match: 22.59% Indels: 41
 DB: 1 Gaps: 16

US-10-007-527A-2 (1-379) x SLMPG33A (1-2661)

QY 3 SerValSerAlaGluHisIleuSerGlyIysAspArgPro-----ProValLeu 18
 Db 586 AATGTTCCCAACCCCTCCCTCTCTCACTCAATCTCCAGCTCAGAGCCGATCCGCC 645
 QY 19 ValSerSerAspIysArgGlyIleArgHisGluLeuArgProIysLeuGlnIleThr 38
 Db 646 GTTTCGTCAAGGCGCGGGAGACGCGCTTCGCCAGCGGGAGTCTCTCGAGCATCAC 705
 QY 39 ThrSerGluThrPheAsnAlaCysGlyIysArgProIleSerGlyValaGly----- 55
 Db 706 AGCTTCCCGCGCTCCGACAGTGCCTGGAGACCAATGAGAGCGCTGATACGCGCTG 765
 QY 56 ---ValThrIleValaAsnGlyProIysSerGlyPheGlyIleuArgSerCysGly 74
 Db 766 GCGGCGAGTGTGTAAACGCG---TCCGTGCTCACTGTGCGGCTTCTGCACTGTGGT 822
 QY 75 LysGlyTyrIleCysProCysCysAlaGlyValaGlyAlaHisArgAlaAspGluIle 94
 Db 823 TCCATCTGGGCTCGCGCGTGTCTCGGCAAGATCAGGGCCAAACCGCGGACAGATC 882
 QY 95 SerGlnValAlaHisGluLeuGlyThrGlySerValAla---MetValThrMetThr 113
 Db 883 GCACAGCGCGCGCCCAAGACCTCCGACGCGGACCGCTGGATGGTCACTGCTGAGC 942
 QY 114 MetArgHisThrAlaGlyIysArgLeuHisAspLeuTyrPheGlyLeuSerAlaIleTyr 133
 Db 943 GCCCGACACAGAACACACACAGAGCTTGAAGCCCTCTTCGACGCGCTGCCAACGCTGG 1002
 QY 134 LysAlaAlaThrAsnGlyIysArgTyrArg-----ThrGlu 145
 Db 1003 CGCAAGCTGTGTCCGCGAGGCTTGGCCGAGAACCCCAAGCGGGAGTCTCGGCGGA 1062
 QY 146 ArgGluMetTyrGlyCysAspGlyTyrValArgAlaValaGluIleThrHisGly---Lys 164
 Db 1063 CGCAGCGCGCTCGAGTCCGCGGCAACATCCGCTCACTCGAAGTCACTTACCGAAGTGT 1122
 QY 165 AsnGlyTyrPheIleValaHisIleAlaLeuMetPheSerGlyAspValSerGluAsn 184
 Db 1123 AACGTTGGCAACCCGACCTTCACTTCTCTCTCGTCAAGAACAGAACACCGAAGA 1182
 QY 185 IleuGluSerPheSerAspAlaMetPheAspArg---TyrThrIysIysValSer 203
 Db 1183 CTC-----GCTACGCAATGACACCGAGGAGCAAACTGAGGCGCTGCTCAAGAA 1236
 QY 204 LeuGlyPheAlaAlaProLeuArgAsnSerGlyIleLeuAspValaArgLysIleGly 223
 Db 1237 GCCGATTGGAACGACGACAGAAACVC-----GACATCACCTGTCCAAAATGC- 1284
 QY 224 GluAlaAspGlnValIleuAlaAlaTyrIleuThrIleAlaSerGly---ValGlyMet 242
 Db 1285 ACGACCCGAGAAAGCGGAGATACATCCGCAAGCTCAAGAGGAAAGTCCGGAAC 1344
 QY 243 GluValGlySerGlyAspGlyIysSerGlyArgHisGlyAsnArgAlaProTyrGlu--- 261
 Db 1345 GAATTCGCCGCGACGACATGAAGCGGTGCTCGGAATCTGGCGCCGTTGGAATG 1404
 QY 262 IleAlaValaAspAlaValaGlyIysAspProGlnAlaLeuGluLeuTyrPargGluPheGlu 281
 Db 1405 CTCGAAATCTTCGCGGACCGGTGACATGGCATGGCTGCCGATTGGCAGAGATACAG 1464
 QY 282 PheGlySerMetGlyArgArgAlaIleAlaThrSerArgGlyLeuArgAlaArg---Ala 300
 Db 1465 AAGGCGACGTTCCGCGCTCGGCGCATACCTGGAGCGTGTCTGCGGAACTCTCTC 1524
 QY 301 GlyLeuGlyAlaGluLeuThrAspAlaGlnIleValaGluGlnGluSerAlaProVal 320
 Db 1525 GGAAGCAGAAAGAGCTCACGACGAGAACTTCAGTGAAGATCGAGAGGAGACG 1584
 QY 321 MetValAlaIleIleProAlaIysSerTyrMetMetIleArgThrCysAlaProTyrVal 340
 Db 1585 TGG---GCATTCGCGCGCGAATCGCTCCGACGATCGCGCT---GTCCCGGCTTC 1638

QY 341 PheGlyGluIleuGlyLeuValGluAlaGly-AlaThrTrp----- 354
 Db 1639 CAGTCCCGCATCTCCGACACCGCCGAGAACCGCGCTTCGAGACTGTCACCTGCTGA 1698
 QY 355 -----GluAsnLeuArgAsp 359
 Db 1699 CTGCCTACCATCTTCTCTACGAGAT 1723

RESULT 11
 SNA243257
 LOCUS SNA243257 9367 bp DNA circular BCT 02-DEC-1999
 DEFINITION Streptomyces natalensis plasmid pSNA1 including spde, kora, spdb2
 and rep genes.
 ACCESSION AJ243257
 VERSION AJ243257.1 GI:6523476
 KEYWORDS KORA protein; rep gene; replication protein; SpdA protein; SPDB2
 protein; transfer protein.
 SOURCE Streptomyces natalensis
 ORGANISM Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 Streptomycineae; Streptomycetaceae; Streptomyces.

REFERENCE
 AUTHORS Mendes, M.V., Aparicio, J.F. and Martin, J.F.
 TITLE Complete nucleotide sequence and characterization of pSNA1 from the
 plasmid-producing Streptomyces natalensis that replicates by a
 rolling circle mechanism
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 9367)
 AUTHORS Mendes, M.V.
 TITLE Direct Submission
 JOURNAL Submitted (23-JUN-1999) Mendes M.V., Microbiology, Inbiotec, AV.
 REAL 1, Leon, 24006, SPAIN

FEATURES
 source
 1. 9367
 /organism="Streptomyces natalensis"
 /mol_type="genomic DNA"
 /db_xref="taxon:68242"
 /plasmid="pSNA1"
 /country="South Africa"
 266..814
 /gene="spda"
 266..814
 /gene="spda"
 /function="putative intramycelial transfer protein"
 /note="ORF4"
 /codon_start=1
 /transl_table=11
 /product="SPDA protein"
 /protein_id="CAB62258.1"
 /db_xref="GI:6523477"
 /translation="MTVIAHPHTGMPSPAGVTYTGNTGGEIPRDADGRRLCEHCLVXP
 VPESLGTKRPSYCSNRQRAYEARQGFVQVTVMALLRSMRANKATQATSRNP
 EAKSRDADQATSRDVAAPATRPAPABRPVPVQPTQALFEEDQADEBAP
 RPARMTDLKOMPRLPGFG"
 complement (1203..1940)
 /gene="kora"
 complement (1203..1940)
 /gene="kora"
 /function="putative transcriptional regulator"
 /note="ORF1"
 /codon_start=1
 /transl_table=11
 /product="KORA protein"
 /protein_id="CAB62259.1"
 /db_xref="GI:6523478"
 /translation="MPKIRSAAPYPMVAVTHRTIMSGELSPGALPTDRGLAENG
 VSRATKALITVLTKEGLLEAVOGSGTRQSAPLHNHNDPAVYRGRIRYAGEY
 ARITASLEPAPADYBAELGIEGAPALIRITVTNASVDALSMSTWIDGLSTAP
 ALVSTRIRREGTWRIETKGVVAVQGVSRVRLATEAQLLSLEPPAAVKLTRTI
 LRDEGILLEYSVTAGRESIYEYTV"
 2096..3025
 /gene="spdb2"
 2096..3025

gene
 CDS

/gene="spdb2"
 /function="putative intramycelial transfer protein"
 /note="ORF2"
 /codon_start=1
 /transl_table=11
 /product="SPDB2 protein"
 /protein_id="CAB62260.1"
 /db_xref="GI:6523479"
 /translation="MPVRPVHRPPAGADPAQLERILRGLTVALVVAAGLAFTATNV
 TOPALDGVSRWIMLMDLPLVASLGTVAIVGGLISEFEVRPGMASILRMPAGLATW
 LNCWGLSIVPTGPPGIPTHLDPAIGLILSVPLVILLALITTRRIVSIAVLD
 EADTCATPDTTPMPAPASAPVPSSEPTILASPPAPAPVPRFORLEVCAGEOVR
 PALPARVSPKPEPAPVPMGADPVALVAVRWMDDELSTQTAERVGLSKTTVA
 NIMRELAHSGORPAGPLPAGQTVLGVA"
 6656..8086
 /gene="rep"
 6656..8086
 /gene="rep"
 /note="ORF3"
 /codon_start=1
 /transl_table=11
 /product="replication protein"
 /protein_id="CAB62261.1"
 /db_xref="GI:6523480"
 /translation="WTDKALSRGCRYALGGVTPKVTBGTAYVLAGLATCGKWHICP
 CCGAKIRSATVELQAAGTAWMDIGINPKIRPAPAAWKDINDTADMAIEAD
 GQAGLGLMTLTMHYSRHTLAELVTQORDAMKALGQAGADKAKDQXGVFV
 RAMEVYGGANGMHPWHVLFEDFPLTPEQDALLEVLYEAMSTLDVGALEPRE
 HGVRLLDLSGHRGPRARYMTKYOQKAAWTTAEMRTDTTQAGDGHRTTETIAVL
 LTEDAADDRATVLMOSTETPAKMRLLVSNGLRKLALVELDPTDGEIALE
 ROSEALVAVLADPMHQAARRKRSILQILKAEGQDQVRLVSWGLVWGVDVLP
 PVDVLEAEQRAEELDRGVGADQORRAHRVAKKSAVPAARKASAAKTAPKKA
 AHCTCGGALDRLAGASGAGHIC"
 BASE COUNT 1391 a 3428 c 3248 g 1300 t
 ORIGIN

Alignment Scores:
 Pred. No.: 3,44e-19 Length: 9367
 Score: 452.50 Matches: 128
 Percent Similarity: 42.82% Conservative: 54
 Best local similarity: 30.12% Mismatches: 158
 Query Match: 22.57% Indels: 85
 Db: 1 Gaps: 14

US-10-007-527A-2 (1-379) x SNA243257 (1-9367)

QY 19 ValSerSerAspLysArgGlyIleArgHisGluLeuArgProLysLeuGlnIleThr 38
 Db 6602 GTGGCGAAGCATCGCGCGATGCGCGTCCGAGACCGGCTGTACGACGAGAG 6661
 QY 39 ThrSerGluThrPheAsnAlaCysGlyArgProIleSerGlyValAsnGlyValThrIle 58
 Db 6662 ACGGACAAAGCGGCTGACGCGCGCGCGGCGGTACGGCTCGGT---GGCGGCTCACCCCC 6718
 QY 59 ValAsnGlyProLysGlySerGlyPhe---GlyGlyLeuArgSerCysGlyLysGlyTyr 77
 Db 6719 AAGGTGACCCCGCGGAAACCGCTACCTCGCGGGCTGGCCACCTCGCGGCAAGGCGAC 6778
 QY 78 IleCysProCysCysAlaGlyLysValGlyAlaHisArgAlaAspGluIleSerGlnVal 97
 Db 6779 ATCTGCCCTGCTGCGCGCGAAGATCGCTCGCGCGGTACGTCGAACTCCAGGCGGCC 6838
 QY 97 ----- 97
 Db 6839 GGGAGCGCTGGGAGAGACATCGCGCGCATCAACGCGCGGAAGATCCGCCCGCGCGCC 6898
 QY 98 -----ValAla 99
 Db 6899 GCCTGAAGAGAGACATCAACGCGCGGAGAGACTGGCGCGCATCGAGAGAGGCC 6958
 QY 100 HisGlnLeuGlyThrGlySerValAlaMetValThrMetThrMetArgHisThrAlaGly 119
 Db 6959 GATGGGCTTCAGAGGCTGTGGCTGCTCGCATGCTGACGCTGACCATCGGCGCATCTCCCGG 7018

QY 120 GlnArgLeuHisAspLeuTrrPThrGlyLeuSerAlaAlaTrpYsaAlaAlaThrAn--- 138
 Db 7019 CACACGCTCGCCGCGCTGTCACCCGACCGCGGACCGCGAGCGAGGAGGAGCGCCCGCGCGAG 7078
 QY 139 -----GlyArgArgTrpArgTrpGluArgGluMetTrpGlyYsaAspGlyTrpValArg 156
 Db 7079 AACCGGGGCGCCGACCTGGCCGAAAGCGAAGAAAGACTACCGGCTTGTGGCTTCGTCGG 7138
 QY 157 AlaValGluLeuThrHisGlyLys---AsnGlyTrpHisValHisValHisAlaLeuLeu 175
 Db 7139 GCGTGGGAGGTGACCTTACCGCGAAGCCAAAGCGTGGGACCCGACCGGATGCTGCTGGTG 7198
 QY 176 MetPheSerGlyAspValSerGluAsnLleLeuGluSerPheSerPheAlaMetPheAsp 195
 Db 7199 TTCTTCGACAGACCGCTGACCCCGGAGCGAGCGAGCGCCCTGAGGAGGATCTACAGG 7258
 QY 196 ArgTrpThrSerLysLeuValSerLeuGlyPheAlaAlaProLeuArgAsnSerGly--- 214
 Db 7259 GCATGGTCCACCGCCCTTCAGGACGTGGGCGCTTACCTGCGGACGCTGAACACGCGCGTC 7318
 QY 215 GlyLeuAspValArgLysLleGlyGlyGluAlaAspGluValLeuAlaAlaTrpLeuThr 234
 Db 7319 CGGCTGACCTGTGCGGGGATGTTGAGGGGCGAGCGCGG-----GCCGCTACCTCAG 7372
 QY 235 LysIleAlaSerGly-----ValGlyMetGluValGlySerGlyAspGly 249
 Db 7373 AAGTACCAAGACCGAGAAAGCGGCGCTGACGACGCGCGCGAGATGACGCGAGCGGACAG 7432
 QY 250 LysSerGlyArgHisGlyAsnArgAlaProTrpGluLleAla-----Val 264
 Db 7433 AAGGCTGGCGCGACCGCCACCGACCGCGCTGTGAATGCGCCGCTCTGCTCACCGAG 7492
 QY 265 AspAlaValAlaGlyLysP---ProGlnAlaLeuGluLeuTrpArgLysPheGluPheGly 283
 Db 7493 GATGCGCGGACGACGACCGCGTGCACGCGCGCTGTGCGGAGTACGAGAGACGAGGAG 7552
 QY 284 SerMetGlyArgArgAlaAlaAlaAlaTrpSerArgGlyLeuArgAlaArgAlaGlyLeuGly 303
 Db 7553 GCGCGGGGATGCGCCCGCTGTCTGCTCAAGCGGCGCTACGCGCGCGCGCTCG 7612
 QY 304 AlaGluLeu-----ThrAspAlaGluLleValGluGluGluGluSerAlaProVal 320
 Db 7613 GTCCAACTCGACACCGCGCACCGGCGGAGATGCGCGTGAAGAACGCCAGGGCGAGGCC 7672
 QY 321 MetValAlaIleLe-----ProAlaArgSer 329
 Db 7673 CTGGGGGTGATCTGTGCGCGACCCCTGCGACGACGACATGCGCCCGCGCAAGCGCGCTCC 7732
 QY 330 TrpMetMetIleArgThrCysAlaProTrpValPheGlyGluLleuGlyLeuValGlu 349
 Db 7733 CTCGACCTGCTCAAGCGCGCGGAGAGGGCGGCGCAGACAAAGTCCGTCCTGCTGGAG 7792
 QY 350 Ala---GlyAlaThrTrpGluAsnLeuArgAspHisLeuHisTrpArgLeuProAlaAla 368
 Db 7793 TCTTGGGGCTGTGTGG-----GGACGC 7816
 QY 369 AspValArgProPro 373
 Db 7817 GAGTCTCTCCGCT 7831
 RESULT 12
 AY172684/c 4603 bp DNA circular BCT 12-JAN-2003
 LOCUS AY172684
 DEFINITION Corynebacterium glutamicum plasmid pAG3, complete sequence.
 ACCESSION AY172684
 VERSION AY172684.1 GI:27657761
 KEYWORDS
 SOURCE Corynebacterium glutamicum
 ORGANISM Corynebacterium glutamicum
 Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 Corynebacterineae; Corynebacteriaceae; Corynebacterium.
 REFERENCE 1 (bases 1 to 4603)

AUTHORS Tauch, A., Puehler, A., Kalinowski, J. and Thierbach, G.
 TITLE Plasmids in Corynebacterium glutamicum and their molecular
 JOURNAL classification by comparative genomics
 AUTHORS Unpublished
 TITLE 2 (bases 1 to 4603)
 JOURNAL Tauch, A., Puehler, A., Kalinowski, J. and Thierbach, G.
 Direct Submission
 Submitted (05-NOV-2002) Department of Genetics, University of
 Bielefeld, Universitaetsstrasse 25, Bielefeld D-33615, Germany
 FEATURES
 source location/Qualifiers
 1. 4603

CDS
 /organism="Corynebacterium glutamicum"
 /mol_type="genomic DNA"
 /strain="22220"
 /db_xref="taxon:1718"
 /plasmid="pAG3"
 complement(200..397)
 /note="orf1"
 /codon_start=1
 /evidence=not_experimental
 /transl_table=1
 /product="hypothetical protein"
 /protein_id="AAO18193.1"
 /db_xref="GI:27657762"
 /translation="MSKQLSPFRAKVLAAVDEVTVPQVAHKIGSTRQKAGSALMAL
 HKANLIDRFRKGLVPLEVDV"
 complement(613..2061)
 /gene="repA"
 complement(613..2061)
 /note="repA"
 /gene="repA"
 complement(613..2061)
 /note="replication protein, essential for pAG3
 replication"
 /codon_start=1
 /transl_table=1
 /product="RepA"
 /protein_id="AAO18194.1"
 /db_xref="GI:27657763"
 /translation="MSLSSRQAEBHATSPAPQATPATTTGRLGTTDTTKHKKVALS
 PALFPAKIMDIQRTWTKYTEDTKTAGCGRWVAGSAVASLVQAQGRARWGLQNSHS
 VWGSPVAVSVCSSRRANVEDALIOAMAGAGLTPAHYGVGVSPPDRRQRGVSNPVV
 ERGTSIMTLTLRHNSKOSTLEVMDAIGCMQAVNTTAMRGAARPAKGRKSGTIAHWY
 RAIEVTHRKKNHNYLHYVLPDRVLSYDEKSLADRFDKMAKAVLWGPASERDR
 GIDVHVAASSDDAKSIGTYCKMLSGIALETTGQVTKAKGDNRPFIQLGLGK
 QYTRKDHALLMEWEGSGRRQVWSQGTQVLTGINELSDQIDISLDNQOSEVYAM
 VGRGAGCAIASDTEKRAVILDAVAATDABGARKKARAVLKLFGVAHTVTALGLAKTD
 LDIVLPHTHDVFKRELIHEAVWSGKQTALPVL"
 2396..2785
 /note="orf3"
 /codon_start=1
 /evidence=not_experimental
 /transl_table=1
 /product="hypothetical protein"
 /protein_id="AAO18195.1"
 /db_xref="GI:27657764"
 /translation="MRSATCLALHALNVRCFVNVNRETLDMTDTVALTRIKCMAT
 KAELEQLKMDQALKEQQLAEFOORNLKPVVTVTRVQYVDEMRKLAAL
 OGWTLQCFANAFAGMIDELSDQEH"
 complement(2840..3323)
 /note="orf4"
 /codon_start=1
 /evidence=not_experimental
 /transl_table=1
 /product="hypothetical protein"
 /protein_id="AAO18196.1"
 /db_xref="GI:27657765"
 /translation="MTRKVRFSANDIPAMIGLVIANENYCTDKQAKTKAKAIAIA
 LNKWDERLGRGKDLAHDSTHIERLRLMELFQKATHDNMDLQVHSLRLOSVR
 IEAGEFTRYGHNVGEVQLAEFDIVE"
 complement(3344..4030)
 /note="orf5"
 /codon_start=1
 /evidence=not_experimental
 /transl_table=1

CDS

```

/product="hypothetical protein"
/protein_id="AA018197.1"
/db_xref="GI:27657766"
/translation="MATPSKPRGLAEQLEKRVSSLELVANVNADHLDSDGKNA
WRTYTAHRTYIAQVEVALETTISMSIDKANDARDALPTATDTSKIVEMELORIT
NRGIPSVQIGYKLVTSHERSPRTALIELEARGLSPEMUGLTKENSPRYAAATE
MMIQHARINSVFCNIRLRTMYKALIDDKKALFNHWNVTBRSADYDMEVPTVFPFKP
TNAETVTRAS"
/complement(4112..4519)
/gene="orf6"
/complement(4112..4519)
/gene="orf6"
/codon_start=1
/evidence_start=1
/evidence_end=1
/product="hypothetical protein"
/protein_id="AA018198.1"
/db_xref="GI:27657767"
/translation="WGRNATPWLVQVSTQIDNISKALNIDMGDHAANASARAEO
OSAKLEADKFRKIALDLSVLKTDCTTGGERRDQIMTAISDKFLXLTSTGAPQ
EVAGQADAAALADRFMYATKAHVHSGSPDNDH"
BASE COUNT      1077 a      1276 c      1112 g      1138 t
ORIGIN

```

```

Alignment Scores:
Pred. No.:      7.84e-18      Length:      4603
Score:          425.50      Matches:      109
Percent Similarity: 49.07%      Conservative: 75
Best Local Similarity: 29.07%      Mismatches: 138
Query Match:      21.22%      Indels:      53
DB:                1          Gaps:      14

```

US-10-007-527a-2 (1-379) x AY172684 (1-4603)

```

QY      2 ThrSerValSerAlaIuhIsleuSerGlyValAspArgProProValIleuValSerSer 21
DB      1971 ACCACTGACACCACTTAAGCACTTAAGGCACTATCACACGCGCTTTTC----- 1918
QY      22 AspLysArgGlyIleArgHisGluLeuArgProLysLeuGlnGlnIleThrThrSerGlu 41
DB      1917 -----CGCGCTAAACTTTGGACCTTACACGCGACACATGTGAAAGTCACAGAGATTAA 1864
QY      42 ThrPheAsnAlaCysGlyArgProIleSerGlyValAsnGlyVal---ThrIleValAsn 60
DB      1863 ACCCTTGGCGGCGCGTGGTGGTGGCGGCGGCTCGCTTCATCGTTCATCGTGGTGG 1804
QY      61 GlyProLysGly---SerGlyPheGlyGlyLeuArgSerCysGlyLysGlyTyrIleCys 79
DB      1803 CAGGACACAGGGGGGTGACGAGTGGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 1744
QY      80 ProCysCysAlaGlyValValGlyAlaHisArgAlaAspGluIleSerGlnValVal--- 98
DB      1743 CCGTGTGACATCGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1684
QY      99 -----AlaHisGlnLeuGlyThrGlySer----- 106
DB      1683 GCGTGGGCTAATGATGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1624
QY      107 -----ValAlaMet 109
DB      1623 AGTGAATCGGAAACAGCGCGCGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1564
QY      110 ValThrMetThrMetArgHisThrAlaGlyGlnArgLeuHisAspLeuThrThrGlyLeu 129
DB      1563 ATGACGTTGACCGCTTAAGCATTAAGCATTAAGCATTAAGCATTAAGCATTAAGCATTA 1504
QY      130 SerAlaAlaIleTyrLysAlaAlaThrAsnGlyArgArgTyrArg----- 143
DB      1503 GCTGCGTGTGGCAGCGCTGTACTTAACACCGCGCGGTGGCGGTGGTGGTGGTGGTGGT 1444
QY      144 ThrGluArgGluMetTyrGlyCysAspGlyTyrValAlaArgAlaValGlnIleThrHisGly 163
DB      1443 GCGGATAAAGTGGCTATGGCATTTGGCCATTTGGGCGGCGGCTATTTAGTGTACCCACG 1384

```

```

QY      164 LysAsnGlyTyrPheIleValHisValHisAlaLeuLeuMetPheSerGlyAspValSerGlu 183
DB      1383 AAGAAATGGCTGGCATGTTTCACTGATCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1324
QY      184 AsnIleLeuGlnSerPheSerAspAlaMetPheAspArgTyrThrSerLysLeuValSer 203
DB      1323 GATGAAAGAAATTCCTTGGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGAT 1264
QY      204 LeuGlyPheAlaAlaProLeuArgAsnSerGlyGlyLeuAspValArgLysIleGlyGly 223
DB      1263 CTGTGATGACGCTGACCGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1207
QY      224 GluAlaAspGlnVal-----LeuAlaIleTyrLeuThrLys---IleAlaSerGlyVal 240
DB      1206 TCGTGTGATGACGCTTAAGCATTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1147
QY      241 GlyMetGluValGlySerGlyAsp---GlyLysSerGlyArgHisGlyLysAsnArgAlaPro 259
DB      1146 GCTGCGGAGACCAACACAGGCGAGGTAACTTAAGAAAGCTTAAGGAGATTAACCGACACT 1087
QY      260 TyrGluIleAlaValAlaAspAlaValGlyGlyAspProGlnAlaLeuGlu-----LeuTyr 277
DB      1086 TTTCAGATTAATCTCGGTGAC---CTGGGAAACAGTACACAAAGCGTGTATGCTTATG 1030
QY      278 ArgGluPheGluPheGlySerMetGlyArgArgAlaIleAlaTyrSerArgGlyLeuArg 297
DB      1029 CTTGATGGTGGAAAGATTTCTTAAGGCGAGCGGACAGCTGTGTGTGTGTGTGTGTGTGTGT 970
QY      298 AlaArgAlaGlyLeuGlyValGluLeuThrAspAlaGlnIleValGlnGln-----Glu 315
DB      969 GATGTTCTTGGATTAAC---GACGTGTGATGATGATGATGATGATGATGATGATGATGAT 913
QY      316 GluSerAlaProValMetValAlaIleLeuProAlaArgSerTyr 330
DB      912 AACGATCACTCCGAAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 868

RESULT 13
E17316      2297 bp      DNA      linear      PART 28-JUL-1999
LOCUS      Bifidobacterium breve gene for replication protein of pNB1.
DEFINITION      E17316
ACCESSION      E17316
VERSION      E17316.1 GI:5711999
KEYWORDS      JP 1998262670-A/1.
SOURCE      Bifidobacterium breve
ORGANISM      Bifidobacterium breve
REFERENCE      Bacteria; Actinobacteria; Actinobacteridae; Bifidobacteriales;
AUTHORS      Bifidobacteriaceae; Bifidobacterium.
TITLES      1 (bases 1 to 2297)
JOURNAL      SHUTTLE VECTOR FOR BIFIDOBACTERIA AND REPLICATION PROTEIN GENE OF
COMMENT      BIFIDOBACTERIA PLASMID
PATENT: JP 1998262670-A 1 06-OCT-1998;
YAKULT HONSHA CO LTD
OS      Bifidobacterium breve
PN      JP 1998262670-A/1
PD      06-OCT-1998
PF      27-MAR-1997 JP 1997091387
PI      IINO TORU, MORISHITA TAKASHI
PC      C12N15/09, (C12N15/09, C12R1:19), (C12N15/09, C12R1:01); CC
strandedness: Double;
CC      topology: linear;
FH      Key      Location/Qualifiers
FT      source      1..2297
FT      /organism='Bifidobacterium breve' FT
FT      /strain='ATCC15698'
FT      /plasmid='pNB1'
FT      -35_signal      387..392
FT      -10_signal      410..415
FT      RBS      525..530
FT      CDS      539..2023
FT      /product='replication protein of pNB1'.

```

```

FEATURES
    source              1..2297
                        /organism="Bifidobacterium breve"
                        /mol_type="genomic DNA"
                        /db_xref="taxon:1685"
BASE COUNT      524 a 659 c 689 g 425 t
ORIGIN
Alignment Scores:
Pred. No.:      1,57e-16      length:      2297
Score:          399.50      Matches:      116
Percent Similarity: 42.25%      Conservative: 64
Best Local Similarity: 27.23%      Mismatches: 169
Query Match:    19.93%      Indels:      77
DB:             6          Gaps:      15

US-10-007-527a-2 (1-379) x E17316 (1-2297)

QY      4 ValSerAlaGluHisLeuSerGly--LysAspArgPro-----15
Db      650 GTAGCCGCGGAGACGGGAGCGGCGCGCTAGCGGCCCGCTTGAGAAAGACGAATMAA 709
QY      16 -----ProValLeuValSerSerAspLysArg-----GlyIleArgHisGlu 29
Db      710 ATTAACCCCTCCCTCCCGCACCGATTTACGGCGCTGCGCTAGCGCTGCGCGCTGAA 769
QY      30 LeuArgProLysLeuGlnGlnIleThrThrSerGlyThrPheAsn-----44
Db      770 AGCCGAAAGATTCGTCGCGCTCATGCCGCGTGGGAGAAAGCTCGGATTCGAGCCGATTAG 829
QY      45 -----AlaCysGlyArgProIleSerGlyValAsnGlyValThrIleVal 59
Db      830 CTGCGCGCGCTGCGACGCTGCGCGCAACCGGTGAC-----ACCGCGGTGGGTGTCATG 883
QY      60 AsnGlyProLysGlySerGlyPheGlyGlyLeuArgSerCysGlyLysGlyTyrIleCys 79
Db      884 ACCAAGCGGAGAAAGCCCGGTTTACAGGCACATGCTGTGCGGCTCGATCTGGCATGG 943
QY      80 ProCysCysAlaGlyLysValGlyAlaHisArgAlaAspGluLysSerGlnValAla 99
Db      944 CCACCTGCTCGGCATCATTCGCCACGAACGGCCGACGAGTGCCTCGGCATCGGG 1003
QY      100 HisGlnLeu-----102
Db      1004 AACCATGCCGAGAACTGAGAAAGCCGCCGACCAATGCGAGCGACGACATGAGCGG 1063
QY      103 -----GlyThrGlySerValAlaMetValThr 111
Db      1064 CAGCGTCTGCCCGGCACTCATGTGTCCGACAGCTTCGAAACATPACATCTTGCGACCT 1123
QY      112 MetThrMetArgHisThrAlaGlyGlnArgLeuHisAspLeuTyrThrGlyLeuSerAla 131
Db      1124 CTCACCTGCGCATGATGCGGACAGATGCCGCTGCCATGACCTTACGCAATCCGCAAG 1183
QY      132 AlaTyrPylAlaAlaHisArgGlyArgTyrThrGluArgGluMetGlyCys 151
Db      1184 GCGTGACGAAAGATGATTAACGAAACCCCTTGCAACGGGCTCGGACCTTGAAATTC 1243
QY      152 AspGlyTyrValArgAlaValGluIleThrHisGlyLysAsnGlyTyrHisValHisVal 171
Db      1244 AGGGGTTTCTCGTCCGCGGATTTGAAATCCTTACGGGTGACGCGCTGCACCTTCACAT 1303
QY      172 HisAlaLeuLeuMetPheSerGlyAspValSerGlyAsnIleLeuGluSerPheSerAsp 191
Db      1304 CATTTGTCATGTTTCTCGATGGCATCTGCGAGATGGGAGGTCAGGCAATGCGACAA 1363
QY      192 AlaMetPheAspArgTyrThrSer-----LysLeuValSerLeuGlyPhe-----206
Db      1364 TGGCTGCTGATGCTGAGAAACCATGTCAAACGGGTTGCGCAGGCAATPACAGAAAAAA 1423
QY      207 -----AlaAlaProLeuArgAsnSerGlyGlyLeuAspValArg---219
Db      1424 GACGGAACCCCTCCTCAACGTCGCTCCG---AAGACAGACACGGCATATGATCTGCAATTC 1480

```

```

QY      220 LysIleGlyGlyGluAlaAspGlnValLeuAlaIleTyrLeuThrLysIleAlaSerGly 239
Db      1481 AAGTCGGGCAAGAGACCCCGGAAACCCCTGCGGCCGAATACATACCAAGATTCAAGCCGAC 1540
QY      240 ValGly-----MetGluValGlySerGlyAspGlyLysSerGlyArgHis 254
Db      1541 AAAGGCGCGCTCCTGCTGCTGCGCAAAATCGCGCGCGGCGGATATCAAGATGCTGTATG 1600
QY      255 GlyAsnArgAlaProTyrPylIleAlaValAlaPheAlaValGly-----AspProGln 272
Db      1601 GGGTCGTTAACCCGTTCCAAATGCTGACCTCCGGGTGCTTCGGGCTGTCATTTCCAG 1660
QY      273 AlaLeuGluLeuTyrArgGluPheGluPheGlySerMetGlyArgArgAlaIleAlaTyr 292
Db      1661 CGGAAAGATCTTGCTGCTCAATACGCGAGCGGACCTTCGCCCGCTGCATACATG 1720
QY      293 SerArgGlyLeuArgAlaArgAlaGlyLeuGlyAlaGluLeuThrAspAlaGlnIleVal 312
Db      1721 TCGCGTGGCTCAGAGAGACATGAGATGTC---GAGAACTGGAAGACGAGAGCTGGCG 1777
QY      313 GluGlnGluGluSerAlaProValMetValAla---IleIleProAlaArgSerTyrMet 331
Db      1778 GGAAGAGCCGAGAACTGCTCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1837
QY      332 MetIleArgThrCysAlaProTyrValPheGlyGluIleLeuGlyLeuValGluAlaGly 351
Db      1838 GACATTCGCAAGAGTGGCTGACGACACTGCGCGCAACGATTGATGCCGCCAAGCC---1894
QY      352 AlaThrTyrPylAsnLeu 357
Db      1895 GAAGACTGGCAGAAATGC 1912

RESULT 14
AF085719/c      5750 bp      DNA      circular BCT 03-JUN-1999
LOCUS      Bifidobacterium breve plasmid pCIB1, complete sequence.
DEFINITION      AF085719
ACCESSION      AF085719
VERSION      AF085719.1 GI:4972586
KEYWORDS
SOURCE
ORGANISM      Bifidobacterium breve
Bifidobacterium breve
Bacteria; Actinobacteria; Actinobacteridae; Bifidobacteriales;
Bifidobacteriaceae; Bifidobacterium.
REFERENCE      1 (bases 1 to 5750)
AUTHORS      O'Riordan K. and Fitzgerald, G.F.
TITLE      Molecular characterisation of a 5.75-kb cryptic plasmid from
Bifidobacterium breve NCFB 2258 and determination of mode of
replication
JOURNAL      FEMS Microbiol. Lett. 174 (2), 285-294 (1999)
MEDLINE      99271179
PUBMED      10339821
AUTHORS      O'Riordan, K. and Fitzgerald, G.F.
TITLE      Direct Submision
JOURNAL      Submitted (22-AUG-1998) Microbiology, University College Cork,
Western Road, Cork, Ireland
FEATURES
    source              1..5750
                        /organism="Bifidobacterium breve"
                        /mol_type="genomic DNA"
                        /strain="NCFB 2258"
                        /db_xref="taxon:1685"
                        /plasmid="pCIB1"
                        /plasmid="pCIB1"
                        /gene="par"
                        /gene="par"
                        /note="partitioning gene"
                        /gene="par"
                        /note="par protein, putative involvement in plasmid
                        partition"
                        /codon_start=1
                        /transl_table=11

```

	gene	/product="partitioning protein" /protein_id="AAD34708.1" /db_xref="GI:4972587" /translation="MILSIASKGGTGTSSIFLACAHTADBSLITLLDAPQGS AIDMEAEKONKDPLPFMYKYAGESDINLKDNBGLTIIDPAGETDLTGTLIKSDI CITTSPTQIDARBAKMTFEAIRGAGGSAIIITRANRTDFROMIDDLANDANN ILFDIVIRGSTYRSKPFCHTPKTYDAADNDELIOKGS" complement(2373..3515) /gene="rep" /note="replication gene" complement(2373..3515) /gene="rep"
CDS		/note="rep; similar to replication proteins of screeptomyces plasmids and other members of the pc194 family that replicate by a rolling circle mechanism" /codon_start=1 /transl_table=1 /product="replication protein" /protein_id="AAD34709.1" /db_xref="GI:4972588" /translation="MTNGEKARFTGYMLCGSIMACPSCAIIRHRAHEVALIGNHA EKLRLAADQWQAHEGQRLPELVSPDSFGNYIEGTLTHDRTMPALMDAILGGA WTMINSGPMFORASERMKIRGFVALEITTVGNMGPHIHVMDLDGDGGEEMO OMLDRMKTWKRYAKAYKKKGNDPNVNAVDENGDLOLPSFGDAGTAARYTTXQ GDKGVITACIELARGDILKRGMSGVNPQLDISGLSDIPQREDIMLETWOTLRK CITSRGLKEDMEVBELDEBELAEKADELPCLVGIVFNRYKDIRSAPETLADALD AAEBRWQEVARLLPGGITLDEEQDALDADEARKGDTLPMSVMV" complement(3926..3934) /note="region conforming to double stranded origin; nicking site of rolling circle plasmids" complement(3985..4845) /gene="f1sk gene homolog" complement(3985..4845) /gene="f1sk gene homolog" /note="hypothetical protein; similar to transfer proteins and f1sk homologs" /codon_start=1 /transl_table=1 /product="unknown" /protein_id="AAD34710.1" /db_xref="GI:4972589" /translation="MKMPMNNTKPEPLRFPRFDPATAGITWGKLANGNDATLYTT NTSVVVGGIGSGSKRTAGMVTVALTSLISGCNTHVIDRGGDWGPFSHATTFFVG DLDTVDHDTLLDDEMCRIASNMQRYSANVNOVKRPPELVIIIDICQSFPNKA GLIGGPKAKVABEITIAATEIIVOKRSGGLFAVIQKPTELSLPALENCENRIC FRVTPEARAVLTGDMDDGPSPTPIDPPARGAIIIGLATGEDVWRFAVSEBEAR AVVNHHGVV"
BASE COUNT	1278 a 1656 c 1616 g 1200 t	
ORIGIN		
Alignment Scores:		
Pred. No.:	1..26e-15	Length: 5750
Score:	392.50	Matches: 106
Percent Similarity:	42.71%	Conservative: 61
Best Local Similarity:	27.11%	Mismatch: 161
Query Match:	19.58%	Indels: 63
DB:	1	Gaps: 12
US-10-007-527A-2 (1-379) x AF065719 (1-5750)		
Oy	25 G Y l le a r h t g s g u l e u a g p r o l y s i e u g i n l e h r h r s e r g u h r P he a s n 44	
Dd	3641 GGTGCGCGCGCTGAAAGCCCAAGATTCCTGTCGTCATGCCGGGTGGCGAAAGCGCTCGGA 3582	
Oy	45 -----Ala c s g l y a r p r o l l e s e r g l y a l a m n 54	
Dd	3581 TTGAGCGCGATTAAAGCTGCCGCGCTGGCGCACGCTGCGCCACACCGCGAGC-----ACC 3528	
Oy	55 G Y v a l t h r l l e a l a s e n g l y P r o l y s e r g l y P h e g l y g l e u a g s e r C y s g l y 74	
Dd	3527 GGCGTGGGTGTCACTGACCAACGCGGAAGAAGCCCGGTTTACAGGCAACCATGCTGTGCGGC 3468	
Oy	75 L y e g l Y T r l l e c y s P r o c y s A l a g l y L y s V a l g l Y a h h s a r g a l a e r g u l l e 94	

LOCUS	DEFINITION	ACCESSION	VERSION
RESUL.T.15			
AF428260	5804 bp	DNA	circular PHG 21-MAR-2002
Proionibacterium phage phiB3, complete genome.			
AF428260			
AF428260.1	GI:19568200		

LOCUS	DEFINITION	ACCESSION	VERSION
AF428260	5804 bp	DNA	circular PHG 21-MAR-2002
Proionibacterium phage phiB3, complete genome.			
AF428260			
AF428260.1	GI:19568200		

Db 4399 GGCATATGTCATGATGCGGGCTATTGCGGCTTCAGCATTCGCAATTCGGTGGCTTGC 4458
QY ProCysCysAlaGlyLysValGlyAlaHisArgAlaAspGluIleSerGlnValAla 99
Db 4459 CCGGTGTGTGCTGCCAGATGCTTCTCAGCGGCAGATGCTGTGACAGTGGCTTGCAGC 4518
QY 100 -----HisGlnLeuGlyThrGlySerValAlaMetValThrMetThrMetArgHisThr 117
Db 4519 AGTGGCATGCGCAGGG-----GGCCGGGCGCATGATGATCAGCTCAGCGTCCGACAGAT 4575
QY 118 AlaGlyGlnArgLeuHisAspLeuTyrThrGlyLeuSerAlaAlaTyrPylsAlaAlaThr 137
Db 4576 CGCAGCGACATCACTCAAAATCGTGTGGAGCGCGCTGCTAAGGGCTGTGCGAAAGCCACG 4635
QY 138 AsnGlyArgArgTyrPyrThrGlnArgGlnMetTyrGlyCysAspGly-----Tyr 154
Db 4636 AATGTCTGCTCTGGATGATGTTCTCGGGTCACTGTTGGGGTGGATGCTGCTGCCCTGG 4695
QY 155 ValArgAlaValGluIleThrHisGlyLysAsnGlyTyrHisValHisValHisAlaLeu 174
Db 4696 CTGGGCTTCGTGGAGTGCAGCGACGCGCATGCCGGTGGCAGCTGCACGTCATGCCCTG 4755
QY 175 LeuMetPheSerGlyAspVal-----SerGlnAsnIleLeuGlnSerPheSerAsp 191
Db 4756 GTCTCTCTGGGCGAGGGGTGCGCAGAGTCCGAGGCGCAGCGGTGGAGCAGATTCCGCCA 4815
QY 192 AlaMetPheAspArgTyrThrPheSerLysLeuValSerLeuGlyPheAlaAlaProLeuArg 211
Db 4816 CGCATGTGGGGTCTTGGAATCGTCCGTGTGTGCTCAGGGTGTGAAGTCAGGGGCTGGAA 4875
QY 212 Asn---SerGlyGlyLeuAspValArgLysIleGlyGlyGlnAlaAspGlnValLeuAla 230
Db 4876 AGCTTGTCAAGGCTCACCTGATCGACCGGATGCCGATGCGCTGGGTGGCTTGCC 4935
QY 231 AlaTyrLeuThrLysIleAlaSerGlyValGlyMetGlnValGlySerGlyAsp---Gly 249
Db 4936 GCTTACTTCTTAAGGCCGCT-----TTGAATTGACTCTTGCCG--GGCAGCAAGGC 4988
QY 250 LysSerGlyArg-----HisGlyAsnArgAlaPro----- 259
Db 4989 CGCTCAGGGCGACAAATTGCACGCCGTTCCAGATCTCGCGCTGCTCACTGGCGACCTCAA 5048
QY 260 TyrGluIleAlaValAspAlaValGly---GlyAspProGlnAlaLeu-GluLeuTyrPyr 278
Db 5049 TGGCGAGGTGGGCTGAGACTGTGCGCAGAGAGCGACCGCGCTGTGGCGCTGTGCA 5108
QY 278 GglnPheGlnPheGlySerMetGlyArgArgAlaIleAlaTyrSerArgGlyLeuArgAl 298
Db 5109 CGAGTGGAGAGAGTCTCGTCCGCTGTCGACAGATGGCGCTCTCCAGGGCCTGTTCGA 5168
QY 298 aaArgAlaGlyLeuGlyAlaGluLeuThr-----AspAlaGlnIleValGluGlnGlu 316
Db 5169 CATGCTCGGTCTGGAATCGTCTGACGCCCTCGAGATGACGAGATTATCGAGACAATGA 5228
QY 316 uSerAlaProValMetValAlaIleIleProAlaArgSerTyrMetMetIleArgThrCy 336
Db 5229 CCTCGATGGGGAGAGACTTTCGTGATCGTCAAGAGAAATGGAACCGCTGTGCGGAC-- 5286
QY 336 saIaProTyrValPheGlyGlnIleLeuGlyLeuValGluAlaGlyAla 352
Db 5287 -----CGTGGCCTGGCTCTCGACGTGCTTGAAGCGCGTTCGT 5322

Search completed: August 22, 2003, 11:37:16
Job time : 3146 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus.p2n model

Run on: August 22, 2003, 09:08:23 ; Search time 264 Seconds

(without alignments)
3875.330 Million cell updates/sec

Title: US-10-007-527A-2

Perfect score: 2005

Sequence: 1 MTSVAHSLSGKDRPPVLVS.....HLHRLPAADVPPPIISVRK 379

Scoring table:

BLOSUM62	
Xgapop 10.0 , Xgapext 0.5	
Ygapop 10.0 , Ygapext 0.5	
Fgapop 6.0 , Fgapext 7.0	
Delop 6.0 , Delext 7.0	

Searched: 2552756 segs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seg length: 0

Maximum DB seg length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+g2n.model -DEV=xlh
-O=/cgm2_1/USPRO.spool/US10007527/runat 20082003 165151 15356/app query.fasta_1.519
-DB=N Geneseq 19Jun03 -OPMT=fastap -SUFFIX=g2n.rng -MINMATCH=0.1 -LOOPT=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.ccd
-LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MOB=LOCAL -OUTPRT=pco -NOM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USRR=US10007527 @CGN 1.1 0 @runat 20082003 165151 15356 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEDBTRY -NEG_SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -MARN TIMEOUT=30 -THREDS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGBEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

N_Geneseq_19Jun03: +
1: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT: *
2: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT: *
3: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT: *
4: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT: *
5: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT: *
6: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT: *
7: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT: *
8: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT: *
9: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT: *
10: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT: *
11: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT: *
12: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT: *
13: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT: *
14: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT: *
15: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT: *
16: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT: *
17: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT: *
18: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT: *
19: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT: *
20: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT: *
21: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT: *
22: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT: *
23: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT: *
24: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT: *
25: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA2003.DAT: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed.

and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2005	100.0	1140	24	ABQ76122
2	2005	100.0	6334	24	Rhodococcus AN12 r
3	2005	100.0	9652	24	Rhodococcus AN12 d
4	2005	100.0	11241	24	plasmid pRRH171 D
5	399.5	19.9	2297	13	plasmid pRRH171 DN
6	337	16.8	9431	13	B. breve essential
7	326	16.3	4447	21	DNA encoding the c
8	326	16.3	4447	21	B. lactofermentum
9	326	16.3	4447	24	B. lactofermentum
10	318	15.9	4447	21	B. lactofermentum
11	318	15.9	4447	21	B. lactofermentum
12	318	15.9	4447	24	B. lactofermentum
13	296	14.8	8500	24	B. lactofermentum
14	282	14.1	4005	24	B. coli/coryneform
15	282	14.1	4005	24	Replicon sequence
16	273	13.6	3741	21	Ketoglucanigenin e
17	273	13.6	3741	21	DNA sequence ID No
18	214	10.7	5648	21	DNA sequence ID No
19	133	6.6	349980	24	Endogenous plasmid
20	127	6.3	3331	10	Bifidobacterium 10
21	125	6.2	349980	24	Sequence of plh1 o
22	125	6.2	349980	24	Bifidobacterium 10
23	125	6.2	349980	24	Bifidobacterium 10
24	125	6.2	349980	24	Bifidobacterium 10
25	122.5	6.1	4411529	22	Bifidobacterium 10
26	122	6.1	2871	23	Bifidobacterium 10
27	120.5	6.0	4403765	22	Bifidobacterium 10
28	119	5.9	1910	19	Pseudomonas aerugi
29	118	5.9	15638	25	Mycobacterium tube
30	117	5.8	68750	21	Nitrosomonas deriv
31	116	5.8	4466	24	S. atroolivaceus 1
32	116	5.8	4466	24	Soraniangium cellulos
33	116	5.8	4466	24	Corynebacteriaceae
34	113	5.6	71989	21	Novel human coding
35	113	5.6	71989	21	S. cellulosum DNA
36	112.5	5.6	349980	24	Soraniangium cellulos
37	112.5	5.6	4403765	22	Bifidobacterium 10
38	112.5	5.6	4411529	22	Bifidobacterium 10
39	112	5.6	4614	22	Soraniangium cellulos
40	112	5.6	349980	24	Mycobacterium tube
41	111.5	5.6	2319	23	Bifidobacterium 10
42	111	5.5	53789	19	DNA encoding novel
43	111	5.5	349980	24	Amycolatopsis medi
44	110.5	5.5	2291	9	Bifidobacterium 10
45	110	5.5	1588	24	Entire amylase gen
					DNA encoding a hum

ALIGNMENTS

RESULT 1

ABQ76122 standard; DNA: 1140 BP.

ABQ76122; 13-JAN-2003 (first entry)

Rhodococcus AN12 replication protein Rep DNA.

KM Plasmid stability protein; replication protein; ethylene forming enzyme;
KM carotenoid biosynthesis enzyme; isoprenoid; pyruvate decarboxylase;
KM polyhydroxyalkanoic acid synthase; PHA synthase; nitrile hydratase;
KM alcohol dehydrogenase; terpene synthase; cholesterol oxidase; gene;
KM shuttle vector; Rep; ds.
XX Rhodococcus erythropolis.
OS

XX MO200255709-A2.
 XX 18-JUL-2002.
 XX 12-DEC-2001; 2001WO-US47868.
 XX 12-DEC-2000; 2000US-254868P.
 XX (DUPO) DU PONT DE NEMOURS & CO E I.
 XX Bramucci MG, Cheng Q, Kostichka KN, Tomb J;
 XX WPI, 2002-557827/59.
 XX P-PSDB; ABB84278.
 XX
 XX New nucleic acid molecule encoding replication protein/plasmid
 XX stability protein, useful in cloning and expression vectors,
 XX particularly shuttle vectors for expression of heterologous genes in
 XX Rhodococcus species -
 XX
 XX Claim 2; Page 63-64; 96pp; English.
 XX
 XX This invention describes a novel nucleic acid encoding a replication
 XX protein or a plasmid stability protein. The product of the invention is
 XX useful for expression of nucleic acid such as genes encoding enzymes
 XX involved in the production of isoprenoid molecules, polyhydroxyalkanoic
 XX acid (PHA) synthases, carotenoid biosynthesis enzymes, nitrile
 XX hydratase, ethylene forming enzyme, pyruvate decarboxylase, alcohol
 XX dehydrogenase, terpene synthases, and cholesterol oxidase in an
 XX Actinomycetales bacteria. The replication protein or plasmid stability
 XX protein are useful in cloning and expression vectors and particularly in
 XX shuttle vectors for the expression of homologous and heterologous genes
 XX in Rhodococcus sp. This sequence encodes the Rhodococcus AN12 Rep
 XX protein described in the disclosure of the invention.
 XX
 XX Sequence 1140 BP, 241 A; 287 C; 347 G; 265 T; 0 other;
 XX
 XX Alignment Scores:
 XX Pred. No.: 2.24e-156 Length: 1140
 XX Score: 2005.00 Matches: 379
 XX Percent Similarity: 100.00% Conservative: 0
 XX Best Local Similarity: 100.00% Mismatches: 0
 XX Query Match: 100.00% Indels: 0
 XX DB: 24 Gaps: 0
 XX
 XX US-10-007-527A-2 (1-379) x ABQ76122 (1-1140)
 QY 1 MethionineSerValSerAlaGluHisLeuSerGlyLysAspArgProProValLeuValSer 20
 Db 1 ATGACCAAGCGTAAGTGTGAAACACCTTTCGCGCAAGAACCAGCTCTCCGTCGTGTCG 60
 QY 21 SerAspLysArgGlyIleArgHisGluLeuArgProLysLeuGlnGlnIleThrThrSer 40
 Db 61 TCCGATTAAGCCGCGGATCCGCAAGACATGCGACCCCAACTTCAACAAATACACACGTCA 120
 QY 41 GluThrPheAsnAlaCysGlyArgProIleSerGlyValAsnGlyValThrIleValAsn 60
 Db 121 GAAACATTAAACGCTGTCGCGGCGGATTTCTGCGCGGAACGATGACCATTTGTCAAC 180
 QY 61 GlyProLysGlySerGlyPheGlyGlyLeuArgSerCysGlyLysGlyTyrIleCysPro 80
 Db 181 GGTCCGAAAGGTTCTGATTCGAGGCGCTTGTTCCTCGCGAAAGGCGTGCATCGCCCC 240
 QY 81 CysCysAlaGlyLysValGlyAlaHisIleArgLysAspGlnIleSerGlnValAlaHis 100
 Db 241 TGCCTGCGGAAAAAGTCGTCACATGTCGACAGCAAAATTTCTCAAGTTGTGCTCAT 300
 QY 101 GlnLeuGlyThrGlySerValAlaMetValThrMetThrMetArgHisThrAlaGlyGln 120
 Db 301 CAACCTCGGACTGCATCTGTTGGCATGTGTGACATGACCAATGACCCCAATACAGCTGCAG 360
 QY 121 ArgLeuHisAspLeuTyrThrGlyLeuSerAlaAlaTyrLysAlaAlaThrAsnGlyArg 140

Db 361 CGGCTCCACGACCTATATGACCTGAGCTTTCGCGACCTCGGAAAGCTGCGCAACAGCTCGT 420
 QY 141 ArgTyrArgThrGluArgGluMetTyrGlyCysAspGlyTyrValAlaGlyAlaValGlnIle 160
 Db 421 CGTTGGCGTTACGAAACGTGAATGTACGCGCTGCGACAGCAATACGTCGCGCTGTGAATC 480
 QY 161 ThrHisGlyLysAsnGlyTyrPheIleValHisIleAlaLeuLeuMetPheSerGlyAsp 180
 Db 481 ACTCAGCGAAAAAAGCGCTGCGACATGTCACAGTTCAACCGCTCACTCACTTCAGTGTGAC 540
 QY 181 ValSerGluSerIleLeuGluSerPheSerAspAlaMetPheAspArgTyrThrSerIys 200
 Db 541 GTGAGTAGAAGAACCTCCGATCTCTTCGGAATGCGATGTCATCGATCGATGCACTTCCAAA 600
 QY 201 LeuValSerLeuGlyPheAlaAlaProLeuArgAsnSerGlyGlyLeuAspValArgLys 220
 Db 601 CTGATATCTCTGGATTTGCTGCGCACCTACGTAAATTCGGGTGTCTCGAGTACGAAG 660
 QY 221 IleGlyGlyLysAlaAspGlnValLeuAlaAlaTyrLeuThrLysIleAlaSerGlyVal 240
 Db 661 ATCGCGGATGAGCTGATCAAGTTCTCGCTGCGATCTGACGAAATTTGCATCTGGCGTT 720
 QY 241 GlyMetGluValGlySerGlyAspGlyLysSerGlyValArgHisGlyAsnArgAlaProTyr 260
 Db 721 GGTATGAGAGTTGTGTAGTGCACGCAAAAAAGTGTGCACATGGCAACCGTGCACCTCG 780
 QY 261 GluIleAlaValAspAlaValGlyValAspProGlnAlaLeuGluLeuTyrPheGluPhe 280
 Db 781 GAAATCCCTGTTGATGAGTGGCGGGGATCCACAAAGCTGTGAACTGTGCGCAAAATTT 840
 QY 281 GluPheGlySerMetGlyArgArgAlaIleAlaTyrSerArgGlyLeuArgAlaArgAla 300
 Db 841 GAGTTTGTTGANTGGACGTCGCGCAATCGCGTGTCCGTGATTCGCGCCGAGCT 900
 QY 301 GlyLeuGlyAlaGluLeuThrAspAlaGlnIleValGlnGlnGlnIleSerAlaProVal 320
 Db 901 GGTCTTGCGGACAGCACTAACAGATGCTCAATCTGTGACGAGAAATCTGCCCGCTC 960
 QY 321 MetValAlaIleIlePheProAlaArgSerThrMetMetIleArgThrCysAlaProTyrVal 340
 Db 961 ATGCTTCGATTCATTCGCGCGCATCGTGAATGATGATTCGATTCGCGCTTACGTC 1020
 QY 341 PheGlyGluIleLeuGlyLeuValGlyAlaGlyValThrTyrPheGluAsnLeuArgAspHis 360
 Db 1021 TTGGCGGAGATTCCTCGGACCTGTCGAGCTGCGCCGACTTGGGAAAATCTTCGATCAC 1080
 QY 361 LeuHisTyrArgLeuProAlaAlaAspValArgProProIleIleSerValArgLys 379
 Db 1081 TTGCATTATCGATTGCGCGCAGCGGATGTGCGGCCCGCATATATATCGTTCCGAAG 1137
 XX
 XX RESULT 2
 XX ABQ76124/c
 XX ID ABQ76124 standard; DNA, 6334 BP.
 XX AC ABQ76124;
 XX DT 13-JAN-2003 (first entry)
 XX XX
 XX Rhodococcus AN12 derived plasmid pAN12 DNA.
 XX
 XX Plasmid stability protein; replication protein; ethylene forming enzyme;
 XX carotenoid biosynthesis enzyme; isoprenoid; pyruvate decarboxylase;
 XX polyhydroxyalkanoic acid synthase; PHA synthase; nitrile hydratase;
 XX alcohol dehydrogenase; terpene synthase; cholesterol oxidase;
 XX shuttle vector; circular; ds.
 XX
 XX Rhodococcus erythropolis.
 XX Synthetic.
 XX
 XX WO200255709-A2.
 XX 18-JUL-2002.

XX 12-DEC-2001; 2001WO-US47868.
PF
XX
XX 12-DEC-2000; 2000US-254868P.
PR
XX
XX (DUPO) DU PONT DE NEMOURS & CO E I.
PA
XX
XX Bramucci MG, Cheng Q, Kostichka KN, Tomb J;
XX WPI; 2002-557827/59.
DR
XX
XX New nucleic acid molecule encoding replication protein/plasmid
PT stability protein, useful in cloning and expression vectors,
PT particularly shuttle vectors for expression of heterologous genes in
PT Rhodococcus species -
XX
XX
XX Claim 19; Page 68-71; 96pp; English.

CC This invention describes a novel nucleic acid encoding a replication
CC protein or a plasmid stability protein. The product of the invention is
CC useful for expression of nucleic acid such as genes encoding enzymes
CC involved in the production of isoprenoid molecules, polyhydroxyalkanoic
CC acid (PHA) synthases, carotenoid biosynthesis enzymes, nitrile
CC hydratases, ethylene forming enzyme, pyruvate decarboxylase, alcohol
CC dehydrogenase, terpene synthases, and cholesterol oxidase in an
CC Actinomycetales bacteria. The replication protein or plasmid stability
CC protein are useful in cloning and expression vectors and particularly in
CC shuttle vectors for the expression of homologous and heterologous genes
CC in Rhodococcus sp. This sequence represents the Rhodococcus AN12 derived
CC plasmid PAN12 DNA described in the disclosure of the invention.

XX
SQ Sequence 6334 BP; 1386 A; 1807 C; 1681 G; 1460 T; 0 other;

Alignment Scores:

Pred. No.:	1,76e-155	Length:	6334
Score:	2005.00	Matches:	379
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	24	Gaps:	0

US-10-007-527a-2 (1-379) x ABQ76124 (1-6334)

QY 1 MetThrSerValSerAlaGluHisLeuSerGlyLysAspArgProProValLeuValSer 20
DB 3051 ATGACGACGCTAAAGTGTGTAACACCTTCCGGGAAACCGGCTCCCGCTCGTCTCG 2992
QY 21 SerAspLysArgGlyLysLeuGHisGluLeuArgProLysLeuGlnGlnIleThrThrSer 40
DB 2991 TCCGATTAAGCGCGCATCCGCGACGAACTCGAACCCAACTTCAACCAACCACTGCA 2932
QY 41 GluThrPheAsnAlaCysGlyArgProIleSerGlyValAsnGlyValThrIleValAsn 60
DB 2931 GAAACATTTAAACGCTGTGTGCGCGCGCATTTCTGGCGTGAACGGTGTGACCATTTGCAAC 2872
QY 61 GlyProLysGlySerGlyPheGlyGlyLeuArgSerCysGlyLysGlyTTPILeCysPro 80
DB 2871 GGTCCGAAAGGTTCTGATTCGGAGGCTTCGTTCCCGCGGAAAGGCTGATTCGCCCC 2812
QY 81 CysCysAlaGlyLysValGlyAlaHisAspAlaAspGluIleSerGlnValAlaHis 100
DB 2811 TGGCTGTGCGGAAAGTCGCTGCACATCGTCAGACGAAATTTCTCAAGTTGTTGCTCAT 2752
QY 101 GluLeuGlyThrGlySerValAlaMetValThrMetThrMetArgHisThrAlaGlyGln 120
DB 2751 CAACCTGGAGACTGATCTGTGTGATGTGATGACATGACATGCGCATACAGCTGTGTGAG 2692
QY 121 ArgLeuHisAspLeuThrThrGlyLysSerAlaAlaThrLysAlaAlaThrAsnGlyArg 140
DB 2691 CGGCTCCAGACCTTAAGACTGACCTTTCGGCAGCGCTCGGAAAGCTCGACCAACGGTCTGT 2632
QY 141 ArgTTPArgThrGluArgGluMetIyGlyCysAspGlyTyrValArgAlaValGluIle 160

DB 2631 CGTTGCGGTACGAAACCTGAATGTACGGCTGCGACGGAATACGTGCGCTGTGTGAATC 2572
QY 161 ThrHisGlyLysAsnGlyTTPHisValHisValHisAlaLeuMetPheSerGlyAsp 180
DB 2571 ACTACGGAAGAAAAAGCTGCGACGTCCACAGTTCACGCGCTACTCATGTTCAGTGTGAC 2512
QY 181 ValSerGluAsnIleLeuGlnSerPheSerAspAlaMetPheAspArgTTPThrSerLys 200
DB 2511 GTGAGTGAGAACATCCCTCGAATCCTTCTCGAATCCGATGTTCCATTCGTGACCTTCCAAA 2452
QY 201 LeuValSerLeuGlyPheAlaAlaProLeuArgAsnSerGlyLysLeuAspValArgLys 220
DB 2451 CTCGTATCTCTGGGATTTGCTGCGCCACTACGTATTCGGGTGTCTCGATGTGCAAG 2392
QY 221 IleGlyGlyGluAlaAspGlnValLeuAlaAlaIleThrThrLysIleAlaSerGlyVal 240
DB 2391 ATCGCGGTGAAGTGTATCAAGTTCGTGCGGTATGTGACGAAATTTGACATCTGGGCTT 2332
QY 241 GlyMetGluValGlySerGlyAspGlyLysSerGlyValArgHisGlyAsnArgAlaProTrp 260
DB 2331 GGTATGAGAGTTGGTGTAGTGTGCGACGAAAGAGTGTGACATGCAACCTGTGACCTTGG 2272
QY 261 GluIleAlaValAspAlaValGlyLysProGlnAlaLeuGlnLeuTrpArgGluPhe 280
DB 2271 GAAATCGCTGTGTATGACGTGGCGGGATCCCAACCGTTGGAACTGTGGCGGAAATTT 2212
QY 281 GluPheGlySerMetGlyArgArgAlaIleAlaThrSerArgGlyLeuAlaArgAla 300
DB 2211 GAGTTTGTTGTGATGGACGTCCGGCATGCGGTGTCCTGGTATGTGCGTGGCGGACT 2152
QY 301 GlyLeuGlyValGluLeuThrAspAlaGlnIleValGluGlnGlnGlnSerAlaProVal 320
DB 2151 GGTCTTGGGCGCAACACTACACATGCTCAATGCTTACGACGAAAGATCTGCCCGGCTC 2092
QY 321 MetValAlaIleIleProAlaArgSerTrpMetMetIleArgThrCysAlaProIyVal 340
DB 2091 ATGTTTGCATTCATTCGCGCGCATCGTGTGATGATGATGTGCACTTGTGGCCTTACGTC 2032
QY 341 PheGlyGluIleLeuGlyLeuValGlyAlaGlyAlaThrTrpGluAsnLeuArgAspHis 360
DB 2031 TTGGCGAGAGTCTCGGACCTGTGGAAGCTGGCGGCGCTTGGGAAATCTTCCGATCAC 1972
QY 361 LeuHisTyrArgLeuProAlaAlaAspValArgProProIleIleSerValArgLys 379
DB 1971 TTGCATTTATGATTTGCCCGGACGCGATGTGCGGCCCCCGAATATATGTTCCGACG 1915
RESULT 3
ID ABQ76126/c
ID ABQ76126 standard; DNA; 9652 BP.
XX
AC ABQ76126;
XX
XX 13-JAN-2003 (first entry)
DB
XX
XX Plasmid pRHB171 DNA.
DE
XX
XX Plasmid stability protein; replication protein; ethylene forming enzyme;
KW carotenoid biosynthesis enzyme; isoprenoid; pyruvate decarboxylase;
KW polyhydroxyalkanoic acid synthase; PHA synthase; nitrile hydratase;
KW alcohol dehydrogenase; terpene synthase; cholesterol oxidase;
KW shuttle vector; circular; de.
XX
OS Synthetic.
XX
XX
XX WO200255709-A2.
XX
PD 18-JUL-2002;
XX
PP 12-DEC-2001; 2001WO-US47868.
XX
XX 12-DEC-2000; 2000US-254868P.
PR
XX
XX (DUPO) DU PONT DE NEMOURS & CO E I.

XX
PI Bramucci MG, Cheng Q, Kostichka KN, Tomb J;
XX
XX WPI; 2002-557827/59.
XX
XX New nucleic acid molecule encoding replication protein/plasmid
PT stability protein, useful in cloning and expression vectors,
PT particularly shuttle vectors for expression of heterologous genes in
PT Rhodococcus species -
XX
XX
XX Claim 27; Page 72; 96pp; English.

XX This invention describes a novel nucleic acid encoding a replication
CC protein or a plasmid stability protein. The product of the invention is
CC useful for expression of nucleic acid such as genes encoding enzymes
CC involved in the production of isoprenoid molecules, polyhydroxyalkanoic
CC acid (PHA) synthases, carotenoid biosynthesis enzymes, nitrile
CC hydratases, ethylene forming enzyme, pyruvate decarboxylase, alcohol
CC dehydrogenase, terpene synthases, and cholesterol oxidase in an
CC Actinomyces bacteria. The replication protein or plasmid stability
CC protein are useful in cloning and expression vectors and particularly in
CC shuttle vectors for the expression of homologous and heterologous genes
CC in Rhodococcus sp. This sequence represents the Plasmid pRHBR171 DNA
CC described in the disclosure of the invention.

XX Sequence 9652 BP; 2097 A; 2729 C; 2559 G; 2267 T; 0 other;

Alignment Scores:

Pred. No.: 2,91e-155 Length: 9652
Score: 2005.00 Matches: 379
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 24 Gaps: 0

US-10-007-527A-2 (1-379) x ABQ76125 (1-9652)

QY 1 MethSerValSerAlaGluHisLeuSerGlyLysAspArgProProValLeuValSer 20
DB 5052 ATGACCGAGCTGAAGTCTGGAACACCTTCCGCGAAAGACCGGCTCCGCTCTGCTGCG 4993
QY 21 SerAspLysArgGlyIleArgHisGluLeuArgProLysLeuGlnIleThrThrSer 40
DB 4992 TCCGATTAAGCGCGGATCCGACAGAACTGCGACCCAACTTCAACAAATCACCACGTC 4933
QY 41 GluThrPheAsnAlaCysGlyArgProLysLeuGlyValAlaGlyValThrIleValAsn 60
DB 4932 GAAACATTTAACGCTGTGCGCGCGGCGGATTTCTGCGCGTGAACGCGTGAACCATTCG 4873
QY 61 GlyProLysGlySerGlyPheGlyGlyLeuArgSerCysGlyLysGlyThrIleCysPro 80
DB 4872 GGTCCGAAAGGTTCTGATTCGAGGCTTGTCTCTGCGAAAGGCTGATTCGCCCC 4813
QY 81 CysCysAlaGlyLysValGlyAlaHisArgHisArgHisArgHisArgHisArgHis 100
DB 4812 TGCCTGCGGAGAAAGTGTGTCGACATGTCGACAGCAAAATTTCTCAAGTTGTCAT 4753
QY 101 GlnLeuGlyThrGlySerValAlaMetValThrMetThrMetAlaGlyHisThrAlaGlyGln 120
DB 4752 CAACCTCGGAGCTGATCTGTTGCGATGTGAAGTGAACCAAGCCCATACAGCTGTCAG 4693
QY 121 ArgLeuHisAspLeuThrThrGlyLeuSerAlaAlaThrPheAlaAlaThrAsnGlyArg 140
DB 4692 CGGCTCCAGCACTTGTGACTGGAATCTTCGCGAGCTTGAAGCTGCGACCAAGCGTCTG 4633
QY 141 ArgTrpArgThrGluArgGluMetTrpGlyCysAspGlyTrpValArgAlaValGluIle 160
DB 4632 CGTTGGCGGTACGGAACGTGAATGTACGCGTGCAGCGATACGCGCGCTGTTGAATC 4573
QY 161 ThrHisGlyLysAsnGlyThrHisValHisValHisAlaLeuLeuMetPheSerGlyAsp 180
DB 4572 ACTCAGGAAAAAAGCGCTGCGACGTCCACGTTCAACGCGTACATGTTCAAGTGTGAC 4513
DB

QY 181 ValSerGluAsnIleLeuGlnSerPheSerAspAlaMetPheAspArgTrpThrSerLys 200
DB 4512 GTGAGTGAGAAACATCTCGAATCTTCTCGAATGCATGTGATCGGTGCAATTCACAA 4453
QY 201 LeuValSerLeuGlyPheAlaAlaProLeuArgAsnSerGlyLysLeuAspValArgLys 220
DB 4452 CTGCTATCTCTCGAATTTGCTGCGCACCACTGAATTCGGTGTGCTCGATGTACAAAG 4393
QY 221 IleGlyGlyGluAlaAspGlnValLeuAlaAlaThrLeuThrValIleAsnSerGlyVal 240
DB 4392 ATCGCGGAGTGAACGTGATCAAGTTCTGCTGCGCTGCTATCTGAAGAAATTCATCTGCGGT 4333
QY 241 GlyMetGluValGlySerGlyAspGlyLysSerGlyArgHisGlyAsnArgAlaProTrp 260
DB 4332 GGTATGAGAGTTGTGTGCTGCGCACGGAAGAGTGTGACATGCAACCTGCACTCGG 4273
QY 261 GluIleAlaValAspAlaValGlyLysAspProGlnAlaLeuGluLeuThrArgGluPhe 280
DB 4272 GAATCGCTGTTGATGACGTGCGCGGAGTCCCAAGCGTTGGAACGTGCGACGAATTT 4213
QY 281 GluPheGlySerMetGlyArgArgAlaIleAlaThrSerArgGlyLeuArgAlaArgAla 300
DB 4212 GAGTTGTGTGATGAGGACGTGCGGACATGCGTGTGCTCGGTGATTTGCGCGAGCT 4153
QY 301 GlyLeuGlyAlaGluLeuThrAspAlaGlnIleValGluGlnGluSerAlaProVal 320
DB 4152 GGTCTTGCGGACAGATTAACAGATGCTCAGATGCTTACAGCAGGAAGATCTGCCCGCTC 4093
QY 321 MetValAlaIleIleProAlaArgSerTrpMetMetIleArgThrCysAlaProTrpVal 340
DB 4092 ATGGTTGCGATCATTTCCGCGCGCATGCTGATGATGATTCGACTTGTGCGCTTACGTC 4033
QY 341 PheGlyGluIleLeuGlyLeuValGluAlaGlyValThrTrpGluLeuArgAspHis 360
DB 4032 TTGCGCGAGATCTCGGACCTGTCGAGCTGCGGCGCATGTTGCGAAATCTTCGATCAC 3973
QY 361 LeuHisTrpArgLeuProAlaAlaAspValArgProProIleIleSerValArgLys 379
DB 3972 TTGATTAATGATTTGCCCGCAGCGAGATGTGCGGCCCGCATTAATATGCTTCGACAG 3916
DB

RESULT 4
ABQ76125/c
ID ABQ76125 standard; DNA; 11241 BP.
XX
AC ABQ76125;
XX
DT 13-JAN-2003 (first entry)
XX
DE Plasmid pRHBR17 DNA.
XX
KM Plasmid stability protein; replication protein; ethylene forming enzyme;
KM carotenoid biosynthesis enzyme; isoprenoid; pyruvate decarboxylase;
KM polyhydroxyalkanoic acid synthase; PHA synthase; nitrile hydratase;
KM alcohol dehydrogenase; terpene synthase; cholesterol oxidase;
KM shuttle vector; circular; ds.
XX
OS Synthetic.
XX
XX W0200255709-A2.
XX
XX 18-JUL-2002.
XX
XX 12-DEC-2001; 2001WO-US47868.
XX
XX 12-DEC-2000; 2000US-254868P.
XX
PA (DUPO) DU PONT DE NEMOURS & CO E I.
XX
PI Bramucci MG, Cheng Q, Kostichka KN, Tomb J;
XX
XX WPI; 2002-557827/59.
XX
XX New nucleic acid molecule encoding replication protein/plasmid
PT

PT stability protein, useful in cloning and expression vectors,
 PT particularly shuttle vectors for expression of heterologous genes in
 Rhodococcus species -

XX Claim 26, Page 71-72; 96pp; English.

XX This invention describes a novel nucleic acid encoding a replication
 CC protein or a plasmid stability protein. The product of the invention is
 CC useful for expression of nucleic acid such as genes encoding enzymes
 CC involved in the production of isoprenoid molecules, polyhydroxyalkanoic
 CC acid (PHA) synthases, carotenoid biosynthesis enzymes, nitrile
 CC hydrolases, ethylene forming enzyme, pyruvate decarboxylase, alcohol
 CC dehydrogenase, terpene synthases, and cholesterol oxidase in an
 CC Actinomycetes bacteria. The replication protein or plasmid stability
 CC protein are useful in cloning and expression vectors and particularly in
 CC shuttle vectors for the expression of homologous and heterologous genes
 CC in Rhodococcus sp. This sequence represents the Plasmid pRHB17 DNA
 CC described in the disclosure of the invention.

SQ Sequence 11241 BP; 2517 A; 3109 C; 2945 G; 2670 T; 0 other;

Alignment Scores:

Pred. No.:	3.se-155	Length:	11241
Score:	2005.00	Matches:	379
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	24	Gaps:	0

US-10-007-527A-2 (1-379) x ABQ76125 (1-11241)

QY 1 MetThrSerValSerIaGluHISLeuSerGlyValAspArgProProValLeuValSer 20
 Db 6641 ATACACGAGCTAAGTGTGAACACCTTTCGCGAAAGACGCGCTCCCGCTCGTGC 6582
 QY 21 SerAspLysArgGlyIleArgHISGluLeuArgProLysLeuGlnIleThrThrSer 40
 Db 6581 TCCGATTAAGCGCGCATCCGGACGAATCCGACCACTTAACAAATCAACAGTCA 6522
 QY 41 GluThrPheAsnAlaCysGlyArgProLysSerGlyValAsnGlyValThrIleValAsn 60
 Db 6521 GAAACATTTAACCCGTGTGGCCGCGCATTTTCGCGTGAACGCTGACCATTTGCAAC 6462
 QY 61 GlyProLysGlySerGlyPheGlyGlyLeuArgSerCysGlyLysGlyTrpIleCysPro 80
 Db 6461 GGTCCGAAAGCTTCGATTCGAGAGCGCTCGTTCTCGCGAAAGCGCTGATCTGCC 6402
 QY 81 CysCysAlaGlyLysValGlyAlaHISArgAlaAspGluLysSerGlnValAlaHIS 100
 Db 6401 TGCCTGCGGGAAGAAAGTCGATCGACATCGTCGACGAAATTTCTCAAGTTGCTCAT 6342
 QY 101 GlnLeuGlyThrGlySerValAlaMetValThrMetThrMetArgHISThrAlaGlyGln 120
 Db 6341 CAACCTCGGAGCTGATCTGTTGCGATGCTGACATGACATGCGCCATACAGCTGCTCAG 6282
 QY 121 ArgLeuHISAspLeuThrThrGlyLeuSerAlaAlaTrpIleValAlaAlaThrAsnGlyArg 140
 Db 6281 CGGCTCCACGACCTTAAGCATGGAATCTTCGACCTTCGCAAACTCGACCAACGATCGT 6222
 QY 141 ArgTrpArgThrGluArgGluMetTrpGlyCysAspGlyTrpValArgAlaValGluIle 160
 Db 6221 CGTTGGCGTACGGAAGCTGAAATGTAAGCTGCGACGATGCTGCGCTGTTGAAATC 6162
 QY 161 ThrHISGlyLysAsnGlyTrpHISValHISValHISAlaLeuLeuMetPheSerGlyAsp 180
 Db 6161 ACTCAGCGAAAAAAGCGCTGCGACGTCACGTTCAACGCGTCACTCATGTTCACTGGTGC 6102
 QY 181 ValSerGluAsnIleLeuGlyLysSerPheSerAspAlaMetPheAspArgTrpThrSerLys 200
 Db 6101 GTGAGTGAAGAAACATCTCGATCTTCTCGATGCGATGCGATGCTGACCTCCAAA 6042
 QY 201 LeuValSerLeuGlyPheAlaAlaProLeuArgAsnSerGlyGlyLeuAspValAlaGlyLys 220

Db 6041 CTGATCTCTGGGATTTGCTGCGCACACGTAACGTAANTCGGGTGTCTCGATGACGAAG 5982
 QY 221 ILeGlyGlyGluAlaAspGlnValLeuAlaAlaTrpLeuThrLysIleAlaSerGlyVal 240
 Db 5981 ATCGCGGTGAAGCTGATCAAGTTCCTGCTGCTATCGACGAAATTCATCTGCGCTT 5922
 QY 241 GlyMetGluValGlySerGlyLysAspGlyLysSerGlyArgHISGlyValAsnArgAlaProTrp 260
 Db 5921 GGTATGGAAGTTGTGATGTCGACGAAAAAGTGTGACATGCGAACCGTGCACCTCTG 5862
 QY 261 GluIleAlaValAspAlaValAlaGlyLysAspProGlnAlaLeuGluLeuTrpArgGluPhe 280
 Db 5861 GAAATGCGCTTGTATCAATGCGCGGCGGATCCACAGCTTGAACCTGTGGCGAATTT 5802
 QY 281 GluPheGlySerMetGlyArgArgAlaIleAlaTrpSerArgGlyLeuArgAlaArgAla 300
 Db 5801 GAGTTGTTGATCATGGAAGTGGCGATCGCGTGTCTCCGTGATTTGGGTGCCAGCT 5742
 QY 301 GlyLeuGlyValGluLeuThrAspAlaGlnIleValGluGlnGluGluSerAlaProVal 320
 Db 5741 GGTCTTGGGCGAAGACTAAACAGATGCTCAGATCGTTGACAGGAAGATCTGCCCGCTC 5682
 QY 321 MetValAlaIleIleProAlaArgSerTrpMetMetIleArgThrCysAlaProTrpVal 340
 Db 5681 ATGGTTGCGATCATTTCCGCGCGATGCTGGAATGATGATTCGACCTTGTGCGCTTTCGTC 5622
 QY 341 PheGlyGluIleLeuGlyLeuValGluAlaGlyAlaThrTrpGluAsnLeuArgAspHIS 360
 Db 5621 TTCGGCGAAGTCTCGACCTCGTCGACGTCGCGACGCTTGGGAAATCTTGTGATTCAC 5562
 QY 361 LeuHISTrpArgLeuProAlaAlaAspValArgProProLysIleSerValArgLys 379
 Db 5561 TTGCATTATCGATGTGCCCGAGCGATGTGCGCCCGCCGATATATATCGGTTGCCAAG 5505

RESULT 5
 AAVS8945
 ID AAVS8945 standard; DNA; 2297 BP.
 XX AC AAVS8945;
 XX DT 05-JAN-1999 (first entry)
 XX XX
 XX DE B. breve essential region gene.
 XX KW Essential region gene; shuttle vector; ds.
 XX OS Bifidobacterium breve.
 XX XX
 FH Key Location/Qualifiers
 FT -35_signal 387..392
 FT -10_signal 410..415
 FT /*tag= a
 FT /*tag= b
 FT RBS 525..530
 FT /*tag= c
 FT CDS 539..2023
 FT /*tag= d
 PN JP10262670-A.
 PD 06-OCT-1998.
 XX 27-MAR-1997; 97JP-0091387.
 XX 27-MAR-1997; 97JP-0091387.
 XX (HONS) YAKULT HONSHA KK.
 XX PA
 DR WPI; 1998-587288/50.
 DR P-PDB; AAM73071.
 XX
 PT Shuttle vector for a Bifidobacterium species - contains specific
 PT elements from existing plasmids e.g. pNBI of ATCC 15698

```

XX PS Disclosure; Page 14-15; 17pp; Japanese.
XX CC This sequence represents the essential region gene from plasmid pNB1
XX CC used in the vector of the invention. The vector is a shuttle vector for a
XX CC Bifidobacterium, and contains: (a) a replicated essential region
XX CC originated from a plasmid pNB1 of Bifidobacterium breve ATCC 15698;
XX CC (b) a replicated essential region originated from a plasmid of E. coli;
XX CC and (c) an antibiotic-resistant gene originated from a plasmid of E. coli
XX CC and an antibiotic-resistant gene functioning by a Bifidobacterium. The
XX CC shuttle vector can be used to transform various Bifidobacteria species.
XX Sequence 2297 BP; 524 A; 659 C; 689 G; 425 T; 0 other;

Alignment Scores:
Pred. No.: 1,78e-23 Length: 2297
Score: 399.50 Matches: 116
Percent Similarity: 42.25% Conservative: 64
Best Local Similarity: 27.23% Mismatches: 169
Query Match: 19.93% Indels: 77
DB: Gaps: 15

US-10-007-527A-2 (1-379) x AAVS8945 (1-2297)

QY 4 ValSerAlaGluHisLeuSerGly---LysAspArgPro----- 15
DB 650 GTAGCCGCGAAGCGGCAAGCGGGCGCGGTAGCGGCCCTTGGAAGAAAGCAATATA 709
QY 16 -----ProValLeuValSerSerAspLysArg-----GlyIleArgHisGlu 29
DB 710 ATACCCCTCTCTCTCCGACCGATTACGGCGCTGCGGTATGATGTCGCGCTGA 769
QY 30 LeuArgProLysLeuGlnGlnIleThrThrSerGluThrPheAsn----- 44
DB 770 AGCGGAAAGATTCTCGCGGTATGCGCGGTGGGAAACGTCGGATTGACGCCATTAG 829
QY 45 -----AlaCysGlyArgProIleSerGlyValAsnGlyValThrIleVal 59
DB 830 CTGCGCGCTGCGCGACGCTGCGGCAACGCGTGAC-----ACCGCGGTGGGTGCATG 883
QY 60 AsnGlyProLysGlySerGlyPheGlyGlyLeuArgSerCysGlyLysGlyTyrPheCys 79
DB 884 ACCAAGGCGGAAAGCGCGGTTCACGACCAATGCTGCGCTGATCTGGGCAATGC 943
QY 80 ProCysCysAlaGlyLysValGlyAlaHisArgAlaAspGluIleSerGlnValAla 99
DB 944 CCCACTGCTCGGCAATCATTCGCAAGAACGCGCCAGAGTCCGCTCGGCATCGGG 1003
QY 100 HisGlnLeu----- 102
DB 1004 AACCATGCGGAAAGCTGAGAAAGCGCGCGGACCAATGCGAGCGAGAACATGAGGG 1063
QY 103 -----GlyThrGlySerValAlaMetValThr 111
DB 1064 CAGCGCTGCTCGCGGCAACTCATGTGTGTCGACAGCTTCGGAATCATCTTCGCGACC 1123
QY 112 MetThrMetArgHisThrAlaGlyGlnArgLeuHisAspLeuTyrThrGlyLeuSerAla 131
DB 1124 CTCACCTGCGCATATGTCGACAAATCCGCTGCCATGACCTTGACGGAATCCCAAG 1183
QY 132 AlaTyrPylAlaAlaThrAsnGlyArgArgTyrArgThrGluArgGluMetCysVal 151
DB 1184 GCGTGAAGCAAGATGATTAACGGAAGCCCTTGCAACGCGGCTCGGAACCTTGAAATTC 1243
QY 152 AspGlyTyrValArgAlaValGluIleThrHisGlyLysAsnGlyTyrPheHisVal 171
DB 1244 AGGGGTTTGTCTCGCGCGGATTGAATATCATCTACGGGTGACCGGCTCATCATTT 1303
QY 172 HisAlaLeuLeuMetPheSerGlyAspValSerGluAsnIleLeuGluSerPheSerAsp 191
DB 1304 CATTTGCTATGTTTTCGATGCGCATCTGACAGATCTGAGAGAGGCTGAGCAATGCAAGCA 1363
QY 192 AlaMetPheAspArgTyrThrSer-----LysLeuValSerLeuGlyPhe----- 206

```

```

DB 1364 TGGCTGCTGATGCTGGAAGAACCATGTCATAGCCCGCTTGCCAAAGCAATACAGAAAAA 1423
QY 207 -----AlaAlaProLeuArgAsnSerGlyGlyLeuAspValArg--- 219
DB 1424 GACGGCAACCCCTACAGCTGCCCG---AACAGCAAGAACGCGATGATGCAATTC 1480
QY 220 TyrIleGlyGlyAlaAlaAspGlnValLeuAlaTyrLeuThrLysIleAlaSerGly 239
DB 1481 AAGTCGGGCAAGACGCGGAACCGCTGCGGCGGAATCATCACCAGATTCAGAGCGAC 1540
QY 240 ValGly-----MetGluValGlySerGlyAspGlyLysSerGlyArgHis 254
DB 1541 AAAGCGGCGTCACTGCTGCTGCAAGAAATCGCGCGCGGATATCAAGATGTCGTATG 1600
QY 255 GlyAsnArgAlaProTyrPylLeuAlaValAlaSerAlaValGly-----AspProGln 272
DB 1601 GGGTGGTTAACCGCTTCAATTCGTCGACTCGGGTGGCTCGGGCTGTCGCAATTTCCAG 1660
QY 273 AlaLeuGluLeuTyrArgGluPheGluPheGlySerMetGlyArgArgAlaIleAlaTyr 292
DB 1661 CCGGAAGATCTGCTGCAATATCTGCGACGCGCACTGCGCGCGCTGCATTAACATG 1720
QY 293 SerArgGlyLeuArgAlaArgAlaGlyLeuGlyAlaGluLeuThrAspAlaGlnIleVal 312
DB 1721 TCGCGGCTTCAAGGAAGACATGAGAGTC---GAGGAATGGAAGACAGAGAGCTGGCG 1777
QY 313 GluGlnGluGluSerAlaProValMetValAla---IleIleProAlaArgSerTyrMet 331
DB 1778 GAGGAAGCGGACGAGACTGCCGCTGCTGCTGCTGCTATGTCGCGCAATGGGTTTACAA 1837
QY 332 MetIleArgThrCysAlaProTyrValPheGlyGluIleLeuGlyLeuValGluAlaGly 351
DB 1838 GACATTCGCAAGAGTGGCGCTGAGACACTGCGCGCACTGATGATGATCGCGCAAGCC 1894
QY 352 AlaThrTyrGluAsnLeu 357
DB 1895 GAAGACTGCGCAGGAAGTC 1912

RESULT 6
AAQ20907/c
ID AAQ20907 standard; DNA; 9431 BP.
XX AC
XX AAQ20907;
XX AC
XX 22-MAY-1992 (first entry)
XX DE
XX CD4; LTR; Streptomyces longisporus; HIV gp120; AIDS; IgG1; T cell;
XX KW Immunoglobulin; surface glycoprotein; virus; MHC class II; ss.
XX OS
XX Homo sapiens.
XX FH
XX Key Location/Qualifiers
XX FT signal_peptide 648..731
XX FT *tag= a
XX FT mat_peptide 732..1286
XX FT /*tag= b
XX FT /product= V1V2
XX FT misc_feature 1287..1331
XX FT /tag= c
XX FT /function= hinge_region
XX FT mat_peptide 1332..1680
XX FT /tag= d
XX FT /product= CH2
XX PN
XX MO9200985-A.
XX XX
XX 23-JAN-1992.
XX PD
XX 01-UNL-1991; 91WO-US04663.
XX PF

```


PS Example; Page 16-18; 28pp; Japanese.

XX The present invention describes a method for transforming a high
 CC temperature-resistant coryneform microbe in which the high temperature-
 CC resistant coryneform microbe is treated with a chemical changing the
 CC structure of cell wall of the microbe and an electric pulse is applied
 CC to a solution containing the microbe having a changed structure of cell
 CC wall and a DNA. Also described are: (1) transforming a high temperature-
 CC resistant coryneform microbe in which the cell wall of a high
 CC temperature-resistant coryneform microbe is weakened and electric pulse
 CC is applied to a solution containing the microbe having weakened cell
 CC wall and a DNA; (2) a transformant of a high temperature-resistant
 CC coryneform microbe in which a recombinant DNA is introduced; and
 CC (3) a transformant of a high temperature-resistant coryneform microbe
 CC carrying both a plasmid vector derived from the high temperature-
 CC resistant coryneform microbe and a plasmid vector derived from
 CC Corynebacterium glutamicum or Brevibacterium lactofermentum. The method
 CC is used for transforming a high temperature-resistant coryneform microbe.
 CC The present sequence encodes a Brevibacterium lactofermentum pSF6
 CC protein sequence, which is used in an example from the present
 CC invention.

XX Sequence 4447 BP; 824 A; 997 C; 1378 G; 1248 T; 0 other;

Alignment Scores:

Pred. No.:	4 6e-17	Length:	4447
Score:	326.00	Matches:	101
Percent Similarity:	47.75%	Conservative:	58
Best Local Similarity:	30.33%	Mismatches:	133
Query Match:	16.26%	Indels:	42
DB:	24	Gaps:	14

US-10-007-527A-2 (1-379) x ABL49733 (1-4447)

QY 34 LeuGInGInleThrThrseryGluThrPheAsnAlaCysGlyArgProIleSeryGlyVal 53
 Db 1396 ATGTACAGATCAACCAATGTAAGCTCTGGCGGGTCCCAAGGTG-GCCAGAGACCA 1454
 QY 54 AaA-----GlyValThrlleVal-AsnGlyProIlySeryGlyPheGlyLeuArg 71
 Db 1455 AACTGTGGCGTGTCTGCTGCTGCTCAACGGT-----GCTTGCAATTTGAAGGCTGTGCA 1508
 QY 71 gserCysGlySeryGlyTrrIleCysProCysGlyValGlyValGlyAlaHisArgAl 91
 Db 1509 AAACCTCTCACTCTCGCTGGGGGTCACTCTGCTGAATTTGAAGCATGGCGCAACGCCG 1568
 QY 91 aaspGluIleSeryGlnValAlaHisGlnLeuGlyThnglySeryValAlaMetValTh 111
 Db 1569 CATTAAGCTGCTATTGCTACTAAGATCACTTGGCGGGTGGCGGCTCATGATGTT 1628
 QY 111 rMet---ThrMetArgHisThrlaGlyGlnArgLeuHisAspLeuTrpThrGlyLeuSe 130
 Db 1629 TGTGGGCACTGTTTCGACCAACCGGTCACAGTCATTGGCGAGTTGAAGCGGATTTAA 1688
 QY 130 rAlaAlaTrpIlyAlaAlaAlaThrsnGlyArgArgTrpArgTrpGluArgGlyMetTrG 150
 Db 1689 GACTGCGTACTCTTCGATGGTGAACCAATCTGAGTGAAGAAACGTCGACGGTACCG 1748
 QY 150 yCysAspGlyTrpValArgAlaValGlnleThrHisGly---LysAsnGlyTrpHisVa 169
 Db 1749 GGTGGAGCACCTTATAGTACTATAGGCTCAAGACTTTGGCGCGCAAGGTTGGACTT 1808
 QY 169 HisValHisAlaLeuLeuMetPheSeryGlyAspValSeryGluAsnleLeuGlnSeryPh 189
 Db 1809 GCACCGCAACATGCTGTTGTTCTTGGATGCTGCACGCTGTGACAGTGAACCTCAAGCGCT 1868
 QY 189 eSeryAlaMetPheAspArgTrpThrserySeryValSeryLeuGlyPheAlaAlaLar 209
 Db 1869 TGAGGATTCATGATTTTCCCGCTGCTGCTGCTGCTGTTAAGCGCGGATGAGACGGCGC 1928
 QY 209 cLeuArgAsnSeryGly---GlyLeuAsp---ValArgylsIleGlyGlyGluAlaAspGl 227
 Db 1929 ACTGGGTGAGCAGCGGGGTCAAACTGATGAGTGTCTACTCTGGGGTGAAGCGCTGCGAA 1988

QY 227 nValleuAlaAlaTrpLeuThrLysIleAlaSeryGlyValGlyMetGluValGlySery 247
 Db 1989 A---ATGGCAACCTTACTCTGCTAAG-----GGCATGTCTCAGAACTGACTGGCTC 2036
 QY 247 yAspGlyYssSeryGlyArgHisGlyAsnArgAlaProTrpGluIleAlaValAspAlaVa 267
 Db 2037 CGCTACTAAACCGCGTCTAAGGGTCTGACACGCGCTTCAGATG---TTGATATGTT 2093
 QY 267 lGlyGlyAspProGlnAlaLeuGlu-----LeuTrpArgGl 279
 Db 2094 GGCAGATCAACACGACCGCGGAGAGATGACCGCTGTTTGGTGGCTCGGTGGCTGA 2153
 QY 279 nPheGluPheGlySeryMetGlyArgArgAlaIleAlaTrpSeryArgGlyLeuArgAlaAr 299
 Db 2154 GTATAGGTGTGTTCTTAAAAACCTCGTTCC---TCCGTGACCGTGGG-----GCTTA 2204
 QY 299 gAlaAlaGlyLeuGlyAlaGluLeuThrAspAlaGlnle-----ValGluGlnGlu- 315
 Db 2205 GCGTCTTGGCGCATGTATACATGACGCTGATGATGCTGCTGGAATGGAAGAAACT 2264
 QY 316 -----GluSeryAlaProValMetValAl 323
 Db 2265 GTACAGCTCGCGGCTCGAAGACCGGAACGGTCCAAATACCCGCTGTGTTGC 2324
 QY 323 allelProAlaArgSeryTrpMetMetIleArgThr 335
 Db 2325 TTTGGTGAAGCCCGATGATGTGAACCTGATTCAGTCT 2361
 RSLT 9
 AAD22582
 ID AAD22582 standard; DNA; 4447 BP.
 AC AAD22582;
 XX 12-FEB-2002 (first entry)
 DT
 DE B. lactofermentum pSF6 replication control region DNA.
 KW Coryneform bacteria; arginine repressor; L-arginine; liver function;
 KM amino acid infusion; de.
 OS Brevibacterium lactofermentum.
 FH Key Location/Qualifiers
 FT CDS 1318..2598
 FT /product= "B. lactofermentum pSF6 replication control
 FT region Protein"
 PN EP1154020-A2.
 PD 14-NOV-2001.
 XX 24-APR-2001; 2001EP-0109457.
 PF 28-APR-2000; 2000JP-0129167.
 PR
 XX (AJIN) AJINOMOTO KK.
 PA Suga M, Asakura Y, Mori Y, Ito H, Kurahashi O;
 PI WPI; 2002-012660/02.
 DR P-PSDB; AAE13530.
 XX
 PT Coryneform bacterium with inactivated arginine repressor, useful for
 PT fermentative production of arginine at high levels -
 PS Example 1; Page 12-15; 30pp; English.
 XX
 CC The present invention relates to Coryneform bacterium in which the
 CC arginine repressor does not function normally and which produces
 CC L-arginine. Coryneform bacterium is used to produce L-arginine which is

CC useful in agents for promoting liver function and in amino acid infusions
CC or in 'comprehensive' amino acid pharmaceutical. Coryneform bacterium
CC accumulate L-arginine to higher levels than wild-type strains, e.g. 120
CC compared with 20, mg/dl. The present sequence is
CC Brevibacterium lactofermentum pSFK6 replication control region DNA.
CC This DNA is used in the construction of shuttle vector for
CC Escherichia coli and Coryneform bacteria and temperature sensitive
CC vector.
XX

SQ Sequence 4447 BP; 824 A; 997 C; 1378 G; 1248 T; 0 other;

Alignment Scores:

Pred. No.:	Length:	4447
Score:	326.00	Matches: 101
Percent Similarity:	47.75%	Conservative: 58
Best Local Similarity:	30.33%	Mismatches: 133
Query Match:	16.26%	Indels: 42
DB:	24	Gaps: 14

US-10-007-527A-2 (1-379) x AAD22582 (1-4447)

```
QY 34 LeuGlnGlnIleThrThrsSerGluThrPheAsnAlaCysGlyAyrProIleSerGlyVal 53
   :::::|||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 1396 ATGTACAGATCCACATATGTAAGGCTCTGGCGGGGTGCCATGAGTG-GCGAGAGGACGA 1454
QY 54 Asn-----GlyValThrIleVal-AsnGlyProIlySerGlyPheGlyIleuAr 71
   ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 1455 AGCTGTGGGAGTCTCTGGTGGTCTGACCGT-----GCTTGCAGATTGAGGGTCTGCA 1508
QY 71 gSerGlyIlySerGlyIlyIleCysProCysGlyAlaGlyIlyValAlaHisArgAl 91
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 1509 AAACCTCTCACTCGCTGGGGGGGTGCACCTGGCTGAATTGGAAGTCATGGCGAAGCGCG 1568
QY 91 aAspGluIleSerGlnValValAlaHisGlnIleGluIlyThrGlySerValAlaMetValm 111
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 1569 CATTGAGCTGGCTATTGCTACTAGATCACTTGGCGGGGTGGCGCCCTCAATGATGTT 1628
QY 111 rMet---ThrMetArgHisThrAlaGlyIlyIlyIleuHisAspLeuThrIlyIleuSe 130
   ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 1629 TGTGGGCACTGTTCGACACACCGCTCACAGTCATTGTGGCAGGTGGAAGCGGATTTAA 1688
QY 130 rAlaAlaTrpIlySAlaAlaThrAsnGlyAyrArgTrpAyrGlnIlyGluMetIlyG 150
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 1689 GACTGGTACTCTTCATGATGATGAAACATCTCAGTGAAGAAAGAACGTGCACGGTACCG 1748
QY 150 yCysAspGlyIlyrValAlaValAlaGluIleThrHisGly---LysAsnGlyTrpHisVa 169
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 1749 GGTGAGACACACTATATGACTATATAGTCAAGACTCTTGGCGCAACGGTTGGACATT 1808
QY 169 HisValHisAlaLeuLeuMetPheSerGlyAspValSerGluAsnIleLeuGluSerPh 189
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 1809 GCACCGCAACATCTCTTTGTTCTTGATCGTCCACTGCTGACGATGAACCTCAAGCGCTT 1868
QY 189 eSerAspAlaMetPheAspArgTrpPheSerIlyLeuValSerIlyPheAlaAlaPr 209
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 1869 TGAAGATTCATGTTTCCCGCTGGTCTGTGTGTTAAGCGCGGATGAGACCGCGCC 1928
QY 209 oIeuAyrAsnSerGly---GlyLeuAsp---ValAyrGlyIleGlyIlyAlaIleAsp 227
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 1929 ACTGCTGAGACACGGGGCTCAAACTTATATAGTGTCTTACTGGGGTGGAGACCTCGCA 1988
QY 227 nValIleuAlaAlaIlyLeuThrIlyrIlyHisIleAserGlyValAlaGlyMetGluValGlySerG 247
   ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 1989 A---ATGGCAACCTACTCGCTAAG-----GGCATGTCTCAGAACTGACGTGGCTC 2036
QY 247 yAspGlyIlySerGlyAyrHisGlyAsnArgAlaPrOTPrGluIleAlaValAlaAspAlaVa 267
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 2037 CGCTACTAAACCGCGCTTAAGGGGTCGTACACGCCGCTTCAGATG---TTGATATGTGT 2093
QY 267 lGlyGlyAspProGlnAlaIleuGlyu-----LeuTrpArgG 279
   ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 2094 GGCCGATCAAGAGCAGCGCCGGCGAGATATGACGCTGTTTGGTGGCTCGATGGCGGTGA 2153
```

```
QY 279 uPheGluPheGlySerMetGlyAyrArgAlaIleAlaTrpSerAyrGlyLeuArgAlaAr 299
   ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 2154 GTATAGAGTGTGGTTCTTAAACCTCGCGTTC---TCTGATGACGTGG-----GCTAA 2204
QY 299 gAlaIlyLeuGlyAlaGluLeuThrAspAlaGlnIle-----ValGluGlnIly-- 315
   ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 2205 GGTGTCTTGGCATTTGATTCATATAGACGTGATGTGTGTGTAAGAAAGAAAGAACT 2264
QY 316 -----GluSerAlaProValMetValAl 323
Db 2265 GTACAGCTCGCCGCTCTGGAAGCACCGGAACGGGTCAATCAACCCGCTTGTGTGC 2324
QY 323 aIleIleProAlaAyrSerTrpMetMetIleAyrThr 335
   ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 2325 TTTGTGTAAGCCCGCATGTATTTGGAATCATTCATTCAGTCT 2361
RESULT 10
AAA90951
ID AAA90951 standard; DNA; 4447 BP.
AC AAA90951;
XX
XX
XX 15-JAN-2001 (first entry)
XX
XX B. lactofermentum pSFK6 coding sequence.
XX
XX Temperature sensitive plasmid; TSRCR; protein production;
XX temperature sensitive replication control region; pSFK6; ds.
XX Brevibacterium lactofermentum.
XX
XX Key location/Qualifiers
XX FT CDS 1318..2601
XX /*tag= a
XX
XX BP1038966-A1.
XX
XX 27-SEP-2000.
XX
XX 16-MAR-2000; 2000BP-0105326.
XX
XX 16-MAR-1999; 990P-0069896.
XX
XX (AJIN ) AJINOMOTO CO INC.
XX
XX Nakamura J, Kanno S, Kimura E, Matsui K, Nakamatsu T;
XX WPI, 2000-573832/54.
XX P-PSDB; AAY97537.
XX
XX Plasmids containing a temperature sensitive replication control regions
XX useful for breeding microorganisms for the production of amino acids by
XX fermentation -
XX
XX Disclosure; Page 19-22; 29pp; English.
XX
XX This sequence encodes the Brevibacterium lactofermentum pSFK6 protein.
XX The invention relates to a plasmid containing a temperature sensitive
XX replication control region (TSRCR) and a marker gene (MG). The TSRCR is
XX derived from plasmid pM330 harboured by Brevibacterium lactofermentum
XX (ATCC 13869) and allows the plasmid to replicate autonomously at a low
XX temperature but does not allow the plasmid to replicate autonomously at
XX an elevated temperature in coryneform bacteria within a temperature
XX range in which the bacteria can grow. The plasmid can be used for
XX modifying a chromosomal gene in a coryneform bacterium, which may be used
XX for the production of useful substances, such as amino acids, by
XX fermentation, to change their genetic traits. Therefore, the plasmid can
XX be used for breeding microorganisms for the production of amino acids by
XX fermentation. The plasmid comprises a TSRCR that allows the plasmid to
XX replicate autonomously at an elevated temperature in coryneform
XX bacteria within a temperature range in which the bacteria can grow.
XX
XX Sequence 4447 BP; 827 A; 993 C; 1375 G; 1252 T; 0 other;
```

Alignment Scores:

Pred. No.: 2.1e-16 Length: 4447
 Score: 318.00 Matches: 100
 Percent Similarity: 47.45% Conservative: 58
 Best Local Similarity: 30.03% Mismatches: 134
 Query Match: 15.86% Indels: 42
 DB: 21 Gaps: 14

US-10-007-527a-2 (1-379) x AAA0951 (1-4447)

```

QY 34 LeuGlnGlnIleThrSerGluThrPheAsnAlaCysGlyArgProIleSerGlyVal 53
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 1396 ATGTACAAGATCAATCAATAGAGCTCTGCGGGGTCATAGGTCG-GCGCGAGAGCA 1454
QY 54 Asn-----GlyValThrIleVal-AsnGlyProGlySerGlyPheGlyIleuArg 71
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 1455 AGCTGTGGCGGTGCTCGGTGCTCAACGCT-----GCTTCCAGATTGAGGCTGTGCA 1508
QY 71 gSerCySGlyLyGlyIleThrIleCysProCySGlyAlaGlyIleValAlaIleArgAl 91
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 1509 AAACCTCACTCGCTGCGGGGTCATCTGCTGATGGAATGGAAGTCATGGCGAGACGCCG 1568
QY 91 aAspGluIleSerGlnValValAlaHisGlnLeuGlyThrGlySerValAlaIleMetValH 111
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 1569 CATTAAGCTGCTATTGCTACTAAGATCACTTGGCGGGGTCGCGCCCTCAATAGATT 1628
QY 111 rMet---ThrMetArgHisThrAlaGlyIleArgLeuHisAspLeuThrThrGlyLeuSe 130
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 1629 TGTGGGACACTGTTGCAACAACCGCTCACAGTCATTTGGCGAGATTGAAGCGGATTAA 1688
QY 130 rAlaAlaIleThrPheAlaAlaThrAsnGlyArgArgThrGluArgGluMetGlyArg 150
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 1689 GACTGGGTACTCTTCATGATGGTGAAGAAACATCTCAGTGAAGAAAGACGTGACAGTACG 1748
QY 150 yCysAspGlyIleValArgAlaValAlaGluIleThrHisGly--LysAsnGlyThrHisVa 169
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 1749 GGTGGGACACACTATATAGTACTATATAGGTCACAGACTCTGGGCGACACGTTGGACATT 1808
QY 169 HisValHisAlaLeuIleuMetPheSerGlyAspValSerGluAsnIleLeuGluSerPh 189
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 1809 GCACCGCAACATGCTTTGTTCTTGATGATGCTGCAGTCTGACAGATGAACTCAAGCATT 1868
QY 189 eSerAspAlaMetPheAspArgTrpThrSerIleuValSerLeuGlyPheAlaIleArg 209
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 1869 TGAAGATTCATATTTCCCGCTGCTGCTGCTGTTAAGCCGCGATGACCGCCG 1928
QY 209 oLeuArgAsnSerGly---GlyLeuAsp---ValArgIleIleGlyIleGluAlaAspG 227
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 1929 ACTGCGTGAAGCACGGGGGTCAAACTTGATCAGGTCTACCTGGGGTGAAGACCTCCGAA 1988
QY 227 nValLeuAlaAlaIleThrIleuThrIleAlaSerGlyValAlaGlyMetGluValGlySerG 247
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 1989 A---ATGGCACTACTCTGGCTTAG-----GGCAGTCTCAAGAACTGACTGGCTC 2036
QY 247 yAspGlyIleSerGlyArgHisGlyAsnArgAlaProIleArgGluIleAlaValaAspAla 267
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 2037 CCTACTAAACCGCGCTCTAAAGGGTCGTAACGCGCTTCAGATG---TTGGAATATGTT 2093
QY 267 lGlyIleAspProGlnAlaIleuGlu-----LeuThrArgG 279
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 2094 GGCGCGATCAAGACGACGGCGCGAGATATGACGCTGTTTGGTGGCTCGGTGGCGGTA 2153
QY 279 uPheGluPheGlySerMetGlyArgArgAlaIleAlaTPSerArgGlyLeuArgAlaArg 299
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 2154 GTATGAGTGTGGTTCTTAAAACTCGCTTCG---TCTTGCTACGTTGGG-----GCTAA 2204
QY 299 gAlaGlyLeuGlyAlaGluLeuThrAspAlaGlnIle-----ValGlnGlnGlu-- 315
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 2205 GCGTGTGGGATGATTATACATAGACGCTGATGACCTCGGAATAGAAAGAAAGAACT 2264
QY 316 -----GluSerAlaProValMetValAl 323
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
  
```

Db 2265 GTACAAGCTCGCCGCTCTGTAAGACACCGGATCGAATCAACCGGTTGCTGTTCG 2324
 QY 323 alleleProAlaArgSerTrpMetIleArgThr 335
 Db 2325 TTGTGTGAAGCCCGATGATTGGAAGAACTGATTCACTGT 2361

RESULT 11

ID ABL49734 standard; DNA; 4447 BP.
 AC ABL49734;
 DT 29-MAY-2002 (first entry)
 DE Brevibacterium lactofermentum p48K nucleotide sequence SEQ ID NO:7.
 KM Kanamycin resistant gene; mutation; high temperature resistance;
 KW coryneform microbe; transformation; gene; ds.
 OS Brevibacterium lactofermentum.
 FH Key Location/Qualifiers
 FT CDS 1318..2601
 FT /tag= a
 FT /product= "p48K"
 PN JP2002017362-A.
 PD 22-JAN-2002.
 PF 05-JUL-2000; 2000JP-0204236.
 PR 05-JUL-2000; 2000JP-0204236.
 PA (AJIN) AJINOMOTO KK.
 PA WPI; 2002-263243/31.
 DR P-PSDB; ABB06342.
 XX Transformation of high temperature-resistant coryneform microbe and
 PT transformant -
 PS Example; Page 20-22; 28pp; Japanese.
 XX The present invention describes a method for transforming a high
 CC temperature-resistant coryneform microbe in which the high temperature-
 CC resistant coryneform microbe is treated with a chemical changing the
 CC structure of cell wall of the microbe and an electric pulse is applied
 CC to a solution containing the microbe having a changed structure of cell
 CC wall and a DNA. Also described are: (1) transforming a high temperature-
 CC resistant coryneform microbe in which the cell wall of a high
 CC temperature-resistant coryneform microbe is weakened and electric pulse
 CC is applied to a solution containing the microbe having weakened cell
 CC wall and a DNA; (2) a transformant of a high temperature-resistant
 CC coryneform microbe in which a recombinant DNA is introduced; and
 CC (3) a transformant of a high temperature-resistant coryneform microbe
 CC carrying both a plasmid vector derived from the high temperature-
 CC resistant coryneform microbe and a plasmid vector derived from
 CC coryneform glutamicum or Brevibacterium lactofermentum. The method
 CC is used for transforming a high temperature-resistant coryneform microbe.
 CC The present sequence encodes a Brevibacterium lactofermentum p48K
 CC protein sequence, which is used in an example from the present
 CC invention.
 SO Sequence 4447 BP; 827 A; 993 C; 1375 G; 1252 T; 0 other;

Alignment Scores:

Pred. No.: 2.1e-16 Length: 4447
 Score: 318.00 Matches: 100
 Percent Similarity: 47.45% Conservative: 58
 Best Local Similarity: 30.03% Mismatches: 134
 Query Match: 15.86% Indels: 42
 DB: 24 Gaps: 14

US-10-007-527a-2 (1-379) x ABL49734 (1-4447)

```

QY 34 LeuGlnGlnIleThrThrsSerGluThrPheAsnAlaCysGlyArgProIleSerGlyVal 53
Db 1396 ATGTACAGATGACCAATGATGAGGCTCGGCGGGGTGCATAGGTG-GCCGAGGAGCA 1454
QY 54 Asn-----GlyValThrIleVal-AsnGlyProIleSerGlyPheGlyGlyLeuAr 71
Db 1455 AGCTGTGGCGGTGTCTGTGCTGTCTACAGGT-----GCTTGCGAGTTTGAAGGCTCTGA 1508
QY 71 gSerCysGlySerGlyTTPileCysProCysAlaGlyValGlyAlaHisArgAl 91
Db 1509 AAACCTCACTCTCGCTGGGGGTCTACTCTGCTGAATGGAAGCATGGCGCAACGCG 1568
QY 91 AAspGluIleSerGlnValValAlaHisGlnLeuGlyThrGlySerValAlaMetValTh 111
Db 1569 CATTGAGCTGGCTAATGCTACTACAGATACATTGGCGGGGTGGCGGCTCATATGTT 1628
QY 111 Met---ThrMetArgHisThrAlaGlyGlnArgLeuHisAspLeuThrGlyLeuSe 130
Db 1629 TGTGGGCACTGTTTCGACCAACCGCTCAACGATTTGGCGAGTTGAAGCGGATTTAA 1688
QY 130 fAlaAlaTrpIlyAlaAlaThrAsnGlyArgArgTrpArgThrGluArgGluMetTyrG 150
Db 1689 GACTCGATCTCTCGATGGTGAACAACTCTCAGTGGAAGAAACGTGACGCGTACGG 1748
QY 150 YCysAspGlyTyrValArgAlaValAlaGluIleThrHisGly--LysAsnGlyTyrHis 169
Db 1749 GGTGAGACACCTTATGAGACTATGAGTCAACAGCTTTGGCGGAACGTTGGCATTT 1808
QY 169 HisValHisAlaLeuLeuMetPheSerGlyAspValSerGluAsnIleLeuGluSerPh 189
Db 1809 GCACCGCAACATGCTGTTGTTCTTGGATGTCACAGTCTGACATGAACTCAAGCATTT 1868
QY 189 eSerAspAlaMetPheAspArgTrpThrSerIlySerValSerLeuGlyPheAlaAlaPr 209
Db 1869 TGAGGATTCATGATTTTCCCGCTGTCTGCTGTGTAAAGCCCGATGAGACGCGCC 1928
QY 209 GluLeuArgAsnSerGly---GlyLeuAsp---ValArgLysIleGlyGlyGluAlaAspG 227
Db 1929 ACTGGCTGAGCACCGGGGTAACTGATAGGTGTTCTTACCTGGGTGAGAGACCTGCCGA 1988
QY 227 nValLeuAlaAlaTyrLeuThrLysIleAlaSerGlyValGlyMetGluValGlySerG 247
Db 1989 A---ATGGCACTTACTCTGCTAAG-----GGCATGCTCAGGAACCTGACGGCTC 2036
QY 247 yAspGlyLysSerGlyArgHisGlyAsnArgAlaProTrpGluIleAlaValAspAlaVa 267
Db 2037 CGCTACTAAACCGCGGTCTAAAGGGGTCTGACACGCGCTTCAGATG---TTGATATGTT 2093
QY 267 lGlyGlyAspProGlnAlaLeuGlu-----LeuTrpArgG 279
Db 2094 GGCCGATCAACGACGCGCGGAGGATATGACGCTGTTTGGTGGCTCGGTGGGTGTA 2153
QY 279 uPheGluPheGlySerMetGlyArgArgAlaIleAlaTrpSerArgGlyLeuArgAlaAr 299
Db 2154 GATATGAGTGTGTTCTTAAAAACCTGCTCG---TCTGTACAGTGGG-----GCTAA 2204
QY 299 gAlaGlyLeuGlyAlaGluLeuThrAspAlaGlnIle-----ValGlnGlnGlu-- 315
Db 2205 GCGTGTCTTGGCATTTGATTACATAGACGCTGATGCTGTGAATGGAAGAAAGAACT 2264
QY 316 -----GluSerAlaProValMetValAl 323
Db 2265 GTACAAAGCTGCGCGGTCTGGAAGACCGGACGCGGTGCAATCAACCGCGTTCGTGCT 2324
QY 323 alleleProAlaArgSerTrpMetMetIleArgThr 335
Db 2325 TTGGTGAAGCCCGAATGATTGGAACATGATTCAGTCT 2361

```

RESULT 12
AAD22583

```

ID AAD22583 standard; DNA; 4447 BP.
XX AC AAD22583;
XX 12-FEB-2002 (first entry)
DE B. lactofermentum p48K replication control region DNA.
XX Coryneform bacteria; arginine repressor; L-arginine; liver function;
XX amino acid infusion; mutant; ds.
OS Brevibacterium lactofermentum.
XX Key Location/Qualifiers
XX CDS 1318..2598
XX FT /tag= a
XX FT /product= "B. lactofermentum pSFK6 replication control
XX FT region Protein"
XX FT replace (1255, C)
XX FT /tag= b
XX FT replace (1534, C)
XX FT /tag= c
XX FT /note= "This mutation causes an amino acid change
XX FT from proline to serine"
XX FT replace (1866, G)
XX FT /tag= d
XX FT replace (2058, G)
XX FT /tag= e
XX FT replace (2187, C)
XX FT /tag= f
XX FT replace (3193, G)
XX FT /tag= g
XX BP154020-A2.
XX 14-NOV-2001.
XX 24-APR-2001; 2001EP-0109457.
XX PR 28-APR-2000; 2000JP-0129167.
XX PA (AJIN ) AJINOMOTO KK.
XX Suga M, Asakura Y, Mori Y, Ito H, Kurahashi O;
XX WPI; 2002-012660/02.
XX DR P-PsDB; AAB13531.
XX PT Coryneform bacterium with inactivated arginine repressor, useful for
XX fermentative production of arginine at high levels -
XX Example 1; Page 17-20; 30pp; English.
XX PS The present invention relates to Coryneform bacterium in which the
XX CC arginine repressor does not function normally and which produces
XX CC L-arginine. Coryneform bacterium is used to produce L-arginine which is
XX CC useful in agents for promoting liver function and in amino acid infusions
XX CC or in 'comprehensive' amino acid pharmaceutical. Coryneform bacterium
XX CC accumulate L-arginine to higher levels than wild-type strains, e.g. 120
XX CC compared with 20 mg/dl. The present sequence is
XX CC Brevibacterium lactofermentum p48K replication control region DNA.
XX CC This DNA is used in the construction of shuttle vector for
XX CC Escherichia coli and Coryneform bacteria and temperature sensitive
XX CC vector.
XX SQ Sequence 4447 BP; 827 A; 993 C; 1375 G; 1252 T; 0 other;

```

Alignment Scores:

Pred. No.:	2,1e-16	Length:	4447
Score:	318.00	Matches:	100
Percent Similarity:	47.45%	Conservative:	58
Best Local Similarity:	30.03%	Mismatches:	134
Query Match:	15.86%	Indels:	42

```

DB:                24          Gaps:                14
US-10-007-527A-2 (1-379) x AAD22583 (1-4447)
QY      34 LeuGlnGlnIleThrThrsSerGluThrPheAsnAlaCysGlyArgProIleSerGlyVal 53
      1396 ATGTACAGATCAACATATAGTAAAGCTCTGCGGGGCGCATAGTG- GCGCAGGACGA 1454
QY      54 Asn-----GlyValThrIleVal-AsnGlyProIleSerGlyPheGlyIleuArg 71
      1455 AGCTGTGGCGGTCCGTCGTCGTCATACGGT-----GCTTCGACGTTTAAAGGCTCGCA 1508
QY      71 gSerCysGlyIleGlyThrIleCysProCysCysAlaGlyIleValAlaIleArgAl 91
      1509 AAACCTCTCACTCGCGGGGGGTCATCTGCTGTAATGGAGTCATGGCGCAACGCCG 1568
QY      91 aaSerGluIleSerGlnValValAlaIleIleGlnGlyThrGlySerValAlaMetValThr 111
      1569 CATTTAGCTGCTGCTATGCTATAGATCACTTGGCGGGGCGGCGCTCATGATGTT 1628
QY      111 rMet---ThrMetArgHisThrAlaGlyIleArgLeuHisAspIleuThrThrGlyLeuSe 130
      1629 TGTGGGCACTGTTCTGCACACACCGCTCACAGTCATTGGCGCAGGTTGAAGCGGATTAA 1688
QY      130 rAlaIleATrPylsAlaAlaIleThrAsnGlyArgArgTrpArgThrGluArgIleuMetYrG 150
      1689 GACTGGGTACTCTTCATGATGGTGAACATCTCATGTGAAGAAAGACGTCCACGCTACG 1748
QY      150 yCysAspGlyTyValAlaArgAlaValAlaGluIleThrHisGly---LysAsnGlyTrpHisVa 169
      1749 GGTGGAGCACACCTATAGTACATATAGGTACAGACTCTGGGCGCAACGCTTGGCACTT 1808
QY      169 HisValHisAlaIleuLeuMetPheSerGlyAspValSerGluAsnIleIleuGluSerP 189
      1809 GCACCGCAACATCTGCTGTTGTTCTTGATGTCATCCAGTGTGACGATGAAGACTCAAGCAT 1868
QY      189 eSerAspAlaMetPheAspArgTrpThrSerIleuValSerIleuGlyPheAlaIleArg 209
      1869 TGAAGATTCATCTTTTCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1928
QY      209 oleuArgAsnSerGly---GlyLeuAsp---ValArgLysIleGlyGluIleAlaIleAsp 227
      1929 ACTGCTGAGACACGCGGTCAACTATGATCAGGTGCTCACTGGGGGTGAGACGCTCGCA 1988
QY      227 nValIleuAlaIleATrIleuThrIleAlaSerGlyValGlyMetGluValGlySerG 247
      1989 A---ATGGCAACTACTGCTGCTAAG-----GGCATGTCTCAGGAACCTGCTGCTC 2036
QY      247 yAspGlyIleSerGlyArgHisGlyAsnArgAlaProTrpGluIleAlaValAspAlaVa 267
      2037 CGCTACTAAACCGCGCTTAAGAGGTCTGACACGCGCTTCAAGATG---TTGATATATGTT 2093
QY      267 lGlyIleAspProGlnAlaIleuGlu-----LeuThrArgG 279
      2094 GCGCGCATCAAGAGGACCGCGCGAGATGACGCTGTTTGTGCTGCTGCTGCTGCTGCTG 2153
QY      279 uPheGluPheGlySerMetGlyArgArgAlaIleAlaIleATrPheArgGlyLeuArgAlaAr 299
      2154 GATGAGGTGGTTGTTCTTAAAAAAGCTGCTG---TCTGTGTACGTTGGG-----GCTAA 2204
QY      299 gAlaGlyLeuGlyAlaGluIleuThrAspAlaGlnIle-----ValGluGlnGlu-- 315
      2205 GCGTGTCTTGGGATGATGATACATAGACGCTGATGTACGTCGTAAGATGGAAGAAAGAACT 2264
QY      316 -----GluSerAlaProValMetValAl 323
      2265 GTACAAAGCTGCGCGGTCTGGAAGACCGCAACCGGTGCGATTCACCCGCGTTGCTGTGC 2324
QY      323 aIleIleProAlaArgSerTrpMetMetIleArgThr 335
      2325 TTTCGTGAAGCCGATGATTGGAACGATTCAGTCT 2361

```

```

ABA93871/C
ID      ABA93871 standard; DNA; 8500 BP.
XX
AC      ABA93871;
XX
DT      02-MAY-2002 (first entry)
XX
DE      E. coli/coryneform bacteria shuttle vector DNA sequence SEQ ID NO:7.
XX
KW      ADH gene; PDC gene; tac promoter; bacteria; shuttle vector; ethanol;
XX      alcohol dehydrogenase; pyruvate decarboxylase; ds.
XX
OS      Escherichia coli.
OS      Coryneform bacterium.
OS      Synthetic.
XX
PN      WO200196573-A1.
XX
PD      20-DEC-2001.
XX
PF      12-JUN-2001; 2001WO-JP04935.
XX
PR      16-JUN-2000; 2000JP-0181625.
XX
PA      (REIN-) RES INST INNOVATIVE TECHNOLOGY EARTH.
XX
PI      Yukawa H;
XX
DR      WPI; 2002-106471/14.
XX
PT      Efficient, high-yielding production of ethanol by recombinant
XX      coryneform bacterium carrying genes expressing pyruvate decarboxylase
XX      and optionally alcohol dehydrogenase activity without substantial
XX      proliferation.
XX
PS      Example 2; Page 24-29; 35pp; Japanese.
XX
CC      The present invention describes a method for producing ethanol using a
XX      coryneform bacterium which is transformed by DNA expressing a pyruvate
XX      decarboxylase activity and if required a gene expressing alcohol
XX      dehydrogenase activity under a regulatory sequence to allow expression
XX      under ethanol production conditions. The bacterium does not proliferate
XX      substantially. Also described are: (1) an expression vector for
XX      transforming the coryneform bacterium which is integrated with a gene
XX      expressing a pyruvate decarboxylase activity and if required a gene
XX      expressing alcohol dehydrogenase activity under a regulatory sequence
XX      to allow the expression for use in the ethanol production; and (2) a
XX      coryneform bacterium transformed with the vector of (1). The method is
XX      used for producing ethanol. The recombinant bacterium can produce
XX      ethanol efficiently with high productivity. The present sequence
XX      represents an E. coli/coryneform bacteria shuttle vector having a PDC
XX      gene linking with a tac promoter and an ADH gene linking with a tac
XX      promoter which is used in an example from the present invention.
XX
SQ      Sequence 8500 BP; 2125 A; 2289 C; 2079 G; 2007 T; 0 other;
XX
Alignment Scores:
Pred. No.:          3e-14          Length:          8500
Score:             296.00          Matches:         105
Percent Similarity: 48.06%          Conservative:     56
Best Local Similarity: 31.34%          Mismatches:      131
Query Match:       14.76%          Indels:          45
DB:                24          Gaps:             14
US-10-007-527A-2 (1-379) x ABA93871 (1-8500)
QY      34 LeuGlnGlnIleThrThrsSerGluThrPheAsnAlaCysGlyArgProIleSerGlyVal 53
      7112 ATGTACAGATCAACATATAGTAAAGCTCTGCGGGGCGCATAGTG- GCGCAGGACGA 7054
QY      54 Asn-----GlyValThrIleVal-AsnGlyProIleSerGlyPheGlyIleuArg 71
      7053 AGCTGTGGCGGTCCGTCGTCGTCATACGGT-----GCTTCGACGTTTAAAGGCTCGCA 7000

```

```

QY 71 gSerCysGlyLysGlyTyrPheCysProCysCysAlaGlyLysValGlyAlaHisArgAl 91
   ::::
Db 6999 AACTCTCACTCTCGCTGGGAGGCTCACTGCTGATGATGAGATGAGGAGGAGCGG 6940
QY 91 aAspGluIleSerGlnValAlaAlaHisGlnLeuGlyThrGlySerValAlaMetValThr 111
   ::::
Db 6939 CATTAGCGCTGATTTGCTACTAATAAGATCACTGGCGGCGGCGGCGCTCATGATGATGT 6880
QY 111 rMet---ThrMetArgHis-ThrAlaGlyGlnArgLeuHisAspLeuTyrThr-GlyLeu 129
   ::::
Db 6879 TGTGGGCACTGTTTCGACACACACCGCTCAAGATCATTTGGCGAGTTGAGAGCGGTATTT 6820
QY 130 SerAlaIleTyrPheSlaAlaIleThrAsnGlyArgArgTyrThrGlnArgGluMetTyr 149
   ::::
Db 6819 AAGACTCGCTACTCTTCATGTGATGTAAGAACATCTCACTGAGAAAGAACAGTCAACGCTAC 6760
QY 150 GlyCysAspGlyTyrValArgAlaValGluIleThrHisGly---LysAsnGlyTyrPhe 168
   ::::
Db 6759 GGCGTGGAGCACACCTAATAGTACATAGAGGTCACAGACTCTTGGCGAAGCGTTGGCAC 6700
QY 169 ValHisValHisAlaLeuLeuMetPheSerGlyAspValSerGluAsnIleLeuGluSer 188
   ::::
Db 6699 TTGCACCG--CACATGTTGTTGTTCTTGATCGTCCACTCTGTCAGACATGAATCTCAAGCGG 6641
QY 189 PheSerAspAlaMetPheAspArgTyrThrSerLysLeuValSerLeuGlyPheAlaAla 208
   ::::
Db 6640 TTGAGGATTCATGTTTCCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 6581
QY 209 ProLeuArgAsnSerGly---GlyLeuAsp---ValArgLysIleGlyGlyGluAlaAsp 226
   ::::
Db 6580 CCACTGCGTGAGACACGGGTCACAACTGATCAAGGTCTCACTCGGCTGAGAACGCTGCG 6521
QY 227 GlnValLeuAlaAlaTyrLeuThrLysIleAlaSerGlyValGlyMetGluValGlySer 246
   ::::
Db 6520 AAA---ATGGCAACCTACCTCGCTAGG-----GGATGCTCTAGAGACCTGACTGGC 6473
QY 247 GlyAspGlyLysSerGlyArgHisGlyAsnArgAlaProTyrGlnIleAlaValAlaAspAla 266
   ::::
Db 6472 TCCGCTACTAAACCGCGCTAAGGGGTGCTGACACCGCTTCAGATG---TTGGATATG 6416
QY 267 ValGlyGlyAspProGlnAlaLeuGlu-----LeuTyrArg 278
   ::::
Db 6415 TTGGCCGATCAAGACACCGCGGAGATATGAGACGTTTGTGCTCGCTGCGCTG 6356
QY 279 GluPheGluPheGlySerMetGlyArgArgAlaIleAlaTyrSerArgGlyLeuArgAla 298
   ::::
Db 6355 GAGTATGAGGTGTGTTCTTAAACCTGCGTTG---TCTGTGCTACGTGGG-----GCT 6305
QY 299 ArgAlaGlyLeuGlyValaGluLeuThrAspAlaGlnIle-----ValGluGlnGlu 315
   ::::
Db 6304 AAGCGTGCCTTGGGCAATGATTCATATAGACCGCTGATGACGTGTAAGTAAGTAAGTAAGTA 6245
QY 316 -----GluSerAlaProValMetVal 322
   ::::
Db 6244 CTGTACAGCTCCCGGCTCTGAAAGCACCGGAAACGGGTGGAATCAACCGCGCTTGTGCTGT 6185
QY 323 AlaIleIleProAlaArgSerTyrMetMetIleArgThr 335
   ::::
Db 6184 GCTTGTGTAAGCCCGATGATTTGAAACATGATTCAGTCT 6146

```

```

RESULT 14
AA17124/c
ID AA17124 standard; DNA; 2401 BP.
XX
XX AA17124;
XX
XX 12-MAR-2002 (first entry)
XX
XX Replicon sequence of Ketogulonigenium endogenous plasmid PADMX6L2.
XX
XX Ketogulonigenium; 2-keto-L-gulonic acid; L-sorbose; sorbitol;
XX
XX plasmid replicon PADMX6L2; replicon; ds.
KW

```

```

XX
OS Ketogulonigenium robustum strain PADMX6L.
XX
XX WO200177348-A2.
XX
XX 18-OCT-2001.
XX
XX 05-APR-2001; 2001WO-US11097.
XX
XX 05-APR-2000; 2000US-194627P.
XX
XX (ARCH) ARCHER-DANIELS MIDLAND CO.
XX
XX (DEUT/) D'ELIA J.
XX
XX (STOD/) STODDARD S F.
XX
XX D'Elia J, Stoddard SF;
XX
XX WPI; 2002-041295/05.
XX
XX New bacterium of Ketogulonigenium genus, useful for producing
XX 2-keto-L-gulonic acid from sorbose or sorbitol, comprises transgene
XX containing DNA sequence from endogenous Ketogulonigenium plasmid -
XX
XX Claim 15; Fig 6; 116pp; English.
XX
XX The present invention relates to a new bacterium of genus
XX Ketogulonigenium. Ketogulonigenium may further comprise a
XX transgene, comprising a dna sequence from an endogenous
XX Ketogulonigenium plasmid. Methods for transforming
XX Ketogulonigenium are also described. The invention is useful for
XX producing 2-keto-L-gulonic acid (2-KLG) from L-sorbose or sorbitol.
XX The present sequence represents the nucleotide sequence of the
XX replicon on the Ketogulonigenium endogenous plasmid PADMX6L2.
XX
SQ Sequence 2401 BP; 497 A; 709 C; 529 G; 666 T; 0 other;

Alignment Scores:
Pred. No.: 9.4e-14 Length: 2401
Score: 282.00 Matches: 86
Percent Similarity: 44.41% Conservative: 41
Best Local Similarity: 30.07% Mismatches: 123
Query Match: 14.06% Indels: 36
DB: 24 Gaps: 14

US-10-007-527a-2 (1-379) x AA17124 (1-2401)
QY 60 AsnGlyProLysGlySerGlyPheGlyGlyLeuArgSerCysGlyTyrPheCys 79
   ::::
Db 1577 GATGGCGGCGCTGGGCACTTCTTCAGGCTTAGACACAGTGCAGATGTGTGGGCTGC 1518
QY 80 ProCysCysAlaGlyLysValGlyAlaHisArgAlaAspGluIleSerGlnValAla 99
   ::::
Db 1517 GCAAGTTGTAGGCGCTCGCATCCGCAATTCGCCCTGCAATGATGACTGTGCTCGG 1458
QY 100 HisGlnLeuGlyThrGlySerVal---AlaMetValThrMetThrArgHisThrAla 118
   ::::
Db 1457 TGGCGCGCGACACACGGGTTTGTGCCGCGTCACTAATCTGACAGCCACACACAAAGCA 1398
QY 119 GlyGlnArgLeuHisAspLeuThrThrGlyLeuSerAlaIleTyrPheSlaAlaHisArg 138
   ::::
Db 1397 GGTGATTCGCTTTTGTGATCTTCTCCAAAATGAGAAAGGCGGTTGGCTGACG 1338
QY 139 GlyArgArgTyrArgThrGlnArgGluMetTyrGlyCysAspGlyTyrValArgAlaVal 158
   ::::
Db 1337 CGCGGAGATGGCGGATTTACATTTGTT-----GGGTCATCATCACTCAACT 1290
QY 159 GluIleThrHisGly---LysAsnGlyTyrPheSlaHisValHisAlaLeu---LeuMet 176
   ::::
Db 1289 GAAATTACTCATAGTATGACAAATGGGTGGGACCGGCACTTTCACGAAATGTGCTGCTA 1230
QY 177 PheSerGlyAspValSerGlyu-----AsnIleLeuGluSerPheSerAspAlaMetPhe 194
   ::::
Db 1229 AAGCTGGGATGAAAGGCAAGCCGCTGATGTGACAGACGTCTAGGCGATCG----- 1176

```

```
QY 135 AsparGTPThrSerLysLeuValSerLeuGlyPheAlaProLeuArgAsnSergly 214
Db 1175 -----TGGCGTGAAGTCTGAAAGGAGTATGTCATGTGGGCGC-----AACGATGCG 1131
QY 215 GlyLeuAspValArgLysIleGlyGlyGluAlaAspGlnValLeuAlaAlaTyrLeuThr 234
Db 1130 GGGTTTGAAGTTCGGGGGCTGCCAATGCGGGCGAC-----TACGTAGCG 1086
QY 235 LysIleAlaSerGlyValGlyMetGluValGlySerglyAspGlyLysSerglyArgHis 254
Db 1085 AAG-----TGGGGCGCTGCCAAGAGTTCGACGCTTTCAGCTCGAAGAGCGGAAACCG 1032
QY 255 GlyAsnArgAlaProTyrGlnIleAlaValAspAlaValGlyLysAspProGlnAlaLeu 274
Db 1031 AAGAGCGACGCCAAGGACGACTACTT-----CAGCGGGCGGATGAC----- 990
QY 275 GluLeuTTPArgGluPheGluPheGlySergMetGly-----ArgArgAlaIleAlaTTP 292
Db 989 GGGCTTTGGCTGCAATATTTCATGCAACGACGCGCAAGCGTACGCGCAATTTGGTGG 930
QY 293 SerArgGlyLeuArgAlaArgAlaGlyLeuGlyAlaGluLeuThrAsp----- 308
Db 929 AGCCAAAGCTTAAAGAAAGATGCGCTTG---GTGAGCTGATGATGACATGAAGCTATG 873
QY 309 AlaGlnIleValGluGlnGluGluSerAlaProValMetValAlaIleIleProAlaArg 328
Db 872 GGTGAAGTTGATGACAGCAGAGCAAGCGGCTGAAATTGTGCGGAGTGAGCAATGAA 813
QY 329 SerTyrMetMetIleArg 334
Db 812 GGTGAAGCAAGATGCGC 795

RESULT 15
AAS17120/C
ID AAS17120 standard; DNA; 4005 BP.
AC AAS17120;
DT 12-MAR-2002 (first entry)
XX
DE Ketogulonigenium endogenous plasmid pADMX6L2 DNA sequence.
XX Ketogulonigenium; 2-keto-L-gulononic acid; L-sorbose; sorbitol;
KW plasmid pADMX6L2; replication protein; cyclic; circular; ds.
XX
OS Ketogulonigenium robustum strain ADMX6L.
XX
Key Location/Qualifiers
FT CDS complement (3960..2560)
FT FT /*tag= a
FT FT /product= "pADMX6L2 replication protein"
FT FT /note= "pADMX6L2 rep ORF"
FT misc_feature 455..2060
FT FT /*tag= b
FT FT /note= "pADMX6L2 replicon, corresponds to AAS17124"
XX
PD MO200177348-A2.
XX
PD 18-OCT-2001.
XX
PF 05-APR-2001; 2001WO-US11097.
XX
PR 05-APR-2000; 2000US-194627P.
XX
PA (ARCH ) ARCHER-DANIELS MIDLAND CO.
PA (DELI/) D'ELIA J.
PA (STOD/) STODDARD S F.
XX
PI D'Elia J, Stoddard SF;
XX
DR WPI, 2002-041295/05.
DR P-PSDB; AAU1039.
```

```
XX
PT New bacterium of Ketogulonigenium genus, useful for producing
PT 2-keto-L-gulononic acid from sorbose or sorbitol, comprises transgene
PT containing DNA sequence from endogenous Ketogulonigenium plasmid -
XX
PS Claim 10; Fig 2; 116pp; English.
XX
CC The present invention relates to a new bacterium of genus
CC Ketogulonigenium. Ketogulonigenium may further comprise a
CC transgene, comprising a DNA sequence from an endogenous
CC Ketogulonigenium plasmid. Methods for transforming
CC Ketogulonigenium are also described. The invention is useful for
CC producing 2-keto-L-gulononic acid (2-KG) from L-sorbose or sorbitol.
CC The present sequence represents Ketogulonigenium endogenous plasmid
CC pADMX6L2 DNA.
CC Note: AAU1039 is encoded by the reverse complement of bases
CC 3960-2560 of this nucleic acid sequence.
XX
SQ Sequence 4005 BP, 866 A, 1174 C, 950 G, 1015 T, 0 other;

Alignment Scores:
Pred. No.: 1.74e-13 Length: 4005
Score: 282.00 Matches: 86
Percent Similarity: 44.41% Conservative: 41
Best Local Similarity: 30.07% Mismatches: 123
Query Match: 14.06% Indels: 36
DB: Gaps: 24

US-10-007-527A-2 (1-379) x AAS17120 (1-4005)
QY 60 AenglyProLysGlySerglyPheGlyGlyLeuArgSerglyLysGlyTTPIleCys 79
Db 3636 GATGCGGGCGCTGCGGCAATTCCTTTTCAGCTCAGACAGTCCGCGAATGTGGGGCTGC 3577
QY 80 ProCysCysAlaGlyLysValGlyAlaHisArgAlaAspGlnIleSergGlnValAla 99
Db 3576 GCAGTTTGTAGGCTGCGATGCGCGCAATTCGCGCTTGTGAATGAATCACTGTGCGG 3517
QY 100 HisGlnLeuGlyThrGlySergVal--AlaMetValThrMetThrMetArgHisThrAla 118
Db 3516 TGGGGCGCGCAACAGGCTTTGTGCGGCTGCTCAATCTTCACGCCAGCAACAGCA 3457
QY 119 GlyGlnArgLeuHisAspLeuTTPThrGlyLeuSerAlaIleTyrLysAlaIleThrAsn 138
Db 3456 GGTGATTCGCTTTTGTGATCTTCTCCAAACATGAAGAGCGCAACAGCGGTTCGTCAG 3397
QY 139 GlyArgArgTTPArgThrGluArgGluMetTyrGlyCysAspGlyTyrValArgAlaVal 158
Db 3396 CGCCGCGAGTGGCGGATTAACCATTTGT-----GGGTCACTCACTCACT 3349
QY 159 GluIleThrHisGly---LysAsnGlyTTPHisValHisValHisAlaLeu---LeuMet 176
Db 3348 GAAATTACTCACTAGCTAATGCAATGGTGGCACCGGCACTTCAAGAAATTCCTTGGTCA 3289
QY 177 PheSerglyAspValSergGlu-----AsnIleLeuGlnSergPheSeraSpAlaMetPhe 194
Db 3288 AGAGCTGGCGATGAAAGCGAGCGCTGATGATGACGAACGCTACGCGATCGC----- 3235
QY 195 AsparGTPThrSerLysLeuValSerLeuGlyPheAlaAlaProLeuArgAsnSergly 214
Db 3234 -----TGGCGTGAAGTCTGAAAGGAGTATGTCATGTGGGCGC-----AACGATGCG 3190
QY 215 GlyLeuAspValArgLysIleGlyGlyGluAlaAspGlnValLeuAlaAlaTyrLeuThr 234
Db 3189 GGGTTTGAAGTTCGGGGGCTGCCAATGCGGGCGAC-----TACGTAGCG 3145
QY 235 LysIleAlaSerGlyValGlyMetGluValGlySerglyAspGlyLysSerglyArgHis 254
Db 3144 AAG-----TGGGGCGCTGCCAAGAGTTCGACGCTTTCAGCTCGAAGAGCGGAAACCG 3091
QY 255 GlyAsnArgAlaProTyrGlnIleAlaValAspAlaValGlyLysAspProGlnAlaLeu 274
Db 3090 AAGAGCGACGCCAAGGACGACTACTT-----CAGCGGGCGGATGAC----- 3049
```

```

Qy 275 GluLeuTPArgGluPheGluPheGlySerMetGly-----ArgArgAlaIleAlaTTrp 292
Db 3048 GGGCTTTGGCTCGAATATTTCATGCAACGAGCGGCAAGCGGCAATTGGTGTGG 2989
Qy 293 SerArgGlyLeuArgAlaArgAlaGlyLeuGlyAlaGluLeuThrAsp----- 308
Db 2988 AGCCCAAGGCTTGAAGAAAGAAATGCGGTTTG--GTCGAGCTGGATGACGATGAAAGCTATG 2932
Qy 309 AlaGluIleValGluGluGluSerAlaProValMetValAlaIleIleProAlaArg 328
Db 2931 GCTGAAGTTGATGCACGACGACGACGCGCCCTTGAATTTGTCCGAGTGGGACAAATGAA 2872
Qy 329 SerTyrMetMetIleArg 334
Db 2871 GGGTGAAGCAAGTGC GC 2854

```

Search completed: August 22, 2003, 10:45:12
 Job time : 307 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Comugen Ltd.

OM protein - nucleic search, using frame_plus.p2n model

Run on: August 22, 2003, 10:35:44 ; Search time 224 Seconds
(without alignments)

3803.813 Million cell updates/sec

Title: US-10-007-527A-2
Perfect score: 2005
Sequence: 1 MTSVAHEHSGKDRPVVLS.....HHYRLPADVAPPIISYRK 379

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Delpop 6.0 , Delext 7.0

Searched: 1517243 seqs, 1124081882 residues

Total number of hits satisfying chosen parameters: 3034486

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-Processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODE=frame+ p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO.spool/US10007527/runat.20082003.165152.15399/app.query.fasta_1.519
-DB=Published Applications NA -OFMT=fastap -SUFFIX=p2n.rnpb -MINMATCH=0.1
-IOOPT=0 -IOOPT=0 -UNITS=bites -START=1 -END=1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALLIGN=200 -THR_SCORE=pct -THR_MAX=100
-THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pct -NORM=ext -HEATSIZE=500 -MINLEN=0
-MAXLEN=2000000000 -USER=US10007527 @CGN 1.1 164 @runat.20082003.165152.15399
-NCPU=6 -ICPU=3 -NO_MMAP -LARGEOUTPUT -NEG_SCORES=0 -WAIT -DSPELACK=100
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FAPOP=6 -FAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications NA:

1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq:*
12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
13: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
14: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
15: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
16: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
17: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	2005	100.0	1140	14	US-10-007-527A-1

2	2005	100.0	1140	14	US-10-007-527A-1	Sequence 1, Appli
3	2005	100.0	6334	14	US-10-007-527A-5	Sequence 5, Appli
4	2005	100.0	6334	14	US-10-007-452-5	Sequence 5, Appli
5	2005	100.0	9652	14	US-10-007-527A-7	Sequence 7, Appli
6	2005	100.0	9652	14	US-10-007-452-7	Sequence 7, Appli
7	2005	100.0	11241	14	US-10-007-527A-6	Sequence 6, Appli
8	2005	100.0	11241	14	US-10-007-452-6	Sequence 6, Appli
9	2005	16.3	4447	9	US-09-835-381-5	Sequence 5, Appli
10	2005	16.3	4447	14	US-10-136-232-5	Sequence 5, Appli
11	2005	15.9	4447	9	US-09-835-381-7	Sequence 7, Appli
12	2005	15.9	4447	14	US-10-196-232-6	Sequence 6, Appli
13	2005	14.1	2401	9	US-09-826-191-6	Sequence 6, Appli
14	2005	14.1	2401	14	US-10-263-666-6	Sequence 6, Appli
15	2005	14.1	2401	14	US-10-261-481-6	Sequence 6, Appli
16	2005	14.1	2401	14	US-10-261-481-6	Sequence 6, Appli
17	2005	14.1	4005	9	US-09-826-191-2	Sequence 2, Appli
18	2005	14.1	4005	14	US-10-263-666-2	Sequence 2, Appli
19	2005	14.1	4005	14	US-10-261-481-2	Sequence 2, Appli
20	2005	14.1	4005	14	US-10-261-481-2	Sequence 2, Appli
21	2005	6.3	9025608	14	US-10-156-761-1	Sequence 1, Appli
22	2005	6.1	2871	9	US-09-815-242-7827	Sequence 7827, Ap
23	2005	5.8	68750	13	US-10-014-717-1	Sequence 1, Appli
24	2005	5.8	1542	9	US-09-833-745-7	Sequence 7, Appli
25	2005	5.6	9025608	14	US-10-156-761-1	Sequence 1, Appli
26	2005	5.5	1536	9	US-09-833-745-12	Sequence 12, Appli
27	2005	5.5	109.5	5.5	2715	Sequence 2245, Ap
28	2005	5.5	13842	10	US-09-861-289-30	Sequence 30, Appli
29	2005	5.3	13842	10	US-09-861-289-30	Sequence 30, Appli
30	2005	5.3	13842	11	US-09-988-384B-30	Sequence 30, Appli
31	2005	5.3	13842	11	US-09-988-384B-30	Sequence 30, Appli
32	2005	5.3	13842	11	US-09-826-191-20	Sequence 30, Appli
33	2005	5.3	36778	10	US-09-861-289-5	Sequence 5, Appli
34	2005	5.3	36778	10	US-09-861-289-5	Sequence 5, Appli
35	2005	5.3	36778	11	US-09-861-289-5	Sequence 5, Appli
36	2005	5.3	37948	11	US-09-988-384B-5	Sequence 5, Appli
37	2005	5.3	38506	11	US-09-793-708-19	Sequence 19, Appli
38	2005	5.3	38506	12	US-10-201-365-1	Sequence 1, Appli
39	2005	5.3	15819	10	US-09-764-877-2165	Sequence 2165, Ap
40	2005	5.3	23229	12	US-10-017-161-2003	Sequence 2003, Ap
41	2005	5.2	3687	11	US-09-917-384-2	Sequence 2, Appli
42	2005	5.2	3687	11	US-09-917-384-2	Sequence 2, Appli
43	2005	5.2	17410	8	US-08-957-425-3	Sequence 3, Appli
44	2005	5.2	1251	14	US-10-156-761-3248	Sequence 3248, Ap
45	2005	5.2	1860	9	US-09-815-242-4156	Sequence 4156, Ap

ALIGNMENTS

RESULT 1
US-10-007-527A-1
; Sequence 1, Application US/10007527A
; Publication No. US20030044807A1
GENERAL INFORMATION:
; APPLICANT: Brannucci, Jean-Francois
; APPLICANT: Brannucci, Michael G.
; APPLICANT: Chen, Qiong
; APPLICANT: Koshchka, Kristy N.
; TITLE OF INVENTION: Rhodococcus Cloning and Expression Vectors
; FILE REFERENCE: C14709 US NA
; CURRENT APPLICATION NUMBER: US/10/007,527A
; PRIOR APPLICATION NUMBER: 2001-12-05
; PRIOR FILING DATE: 2000-12-12
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 1
; LENGTH: 1140
; TYPE: DNA
; ORGANISM: Rhodococcus AM12
US-10-007-527A-1
Alignment Scores: 1.08e-213 Length: 1140
Pred. No.: 1

Score: 2005.00 Matches: 379
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 14 Gaps: 0

US-10-007-527A-2 (1-379) x US-10-007-527A-1 (1-1140)

QY 1 MetThrSerValSerAlaGluHisLeuSerGlyLysAspArgProProValLeuValSer 20
 Db 1 ATGACCAGCGTAAAGTCTTAACACCTTCCGGCAAGACCGGCTCCCTCGTGTG 60
 QY 21 SerAspLysArgGlyTLeaRgHisLeuEuArgProLysLeuGlnGlnLeThrThrSer 40
 Db 61 TCCGATTAAGCGCGCATCCGGACAGAACCTGCCAACCAATTCAACAATCACCAGTCA 120
 QY 41 GluThrPheAsnAlaCysGlyArgProLieserGlyValAsnGlyValThrIleValAsn 60
 Db 121 GAAACATTTAAAGCCCTGTGCGCGCCGATTTCTGGCGTAACGGGTGACCATTTGTCAAC 180
 QY 61 GlyProLysGlySerGlyPheGlyGlyLeuArgSerCysGlyLysGlyTrrPileCysPro 80
 Db 181 GGTTCGAAAGTTCCTGAGATTCGAGAGCCCTTCGTTCTGCGGAAAGGCTGATCTGCCCC 240
 QY 81 CysCysAlaGlyLysValGlyAlaHisArgAlaAspGluLieserGlnValAlaHis 100
 Db 241 TCGTGTGCGGAAAAAGTCGGTCACATCGTGACAGAAATTTCTCAAGTTGTGCTCAT 300
 QY 101 GlnLeuGlyThrGlySerValAlaMetValThrMetMetArgHisThrAlaGlyGln 120
 Db 301 CAACCTGGAGACTGGATCTGTGCGATGTGACGATGACCATCGCCATACAGCTGATCAG 360
 QY 121 ArgLeuHisAspLeuThrPheGlyLeuSerAlaAlaTrrPylsAlaAlaThrAsnGlyArg 140
 Db 361 CGGCTCCAGCACTATGACCTGACCTTTCGGCAGCCTGAAACCTGCCACCAACGCTCT 420
 QY 141 ArgTrrPArgThrGluArgGluMetTyrGlyCysAspGlyTyrValArgAlaValGluIle 160
 Db 421 CGTTGGCGTACGGAAGTGAATGTACGCGCTGCGACGAGTACGTCGCGCTGTGTAATC 480
 QY 161 ThrHisGlyLysAsnGlyTrrPHisValHisValHisAlaLeuLeuMetPheSerGlyAsp 180
 Db 481 ACTCAACGAAAAAAGCGCTGGCAGCGTCCAGCTTCAAGCCCTCACTGTCAGTGTGAC 540
 QY 181 ValSerGluAsnIleLeuGlnSerPheSerAspAlaMetPheAspArgTrrPThrSerLys 200
 Db 541 GAGAGGAGAAACATCTCTGATCTCTCTCGAGTGCATGTCGATCGGACCTTCCAA 600
 QY 201 LeuValSerLeuGlyPheAlaAlaProLeuArgAsnSerGlyLysLeuAspValArgLys 220
 Db 601 CTCGTATCTCTGGGATTTGCTGGCGCACATACGTAATTCGGGTGTCTCGATGTACAAAG 660
 QY 221 IleGlyGlyGluAlaAspGlnValLeuAlaAlaTyrLeuThrLysIleAlaSerGlyVal 240
 Db 661 ATCGGGGGTGAAGCTATCAAGTTCGCGCGATCTGACGAAATTCGATCTGGCGGT 720
 QY 241 GlyMetGluValGlySerGlyAspGlyLysSerGlyArgHisGlyLysAsnArgAlaProTrr 260
 Db 721 GGTATGAGAGTTGTGTGTGGCGACGGAAGAAAGTGTGTGCAATGTGCAACCTGTGACCT 780
 QY 261 GluIleAlaValAspAlaValGlyLysAspProGlnAlaLeuGlnLeuTrrPArgGluPhe 280
 Db 781 GAAATCGCTGTATCATCAGTGGCGCGGAGTCCCAAGCCGTGGACCTGTGGCGAGATTT 840
 QY 281 GluPheGlySerMetCysArgArgAlaIleAlaTrrSerArgGlyLeuArgAlaAspGly 300
 Db 841 GAGTTTGTGTGATGGAGACGTCGGGCAATGCGCTGTGTCCGTGTGATGTGGTCCGAGCT 900
 QY 301 GlyLeuGlyAlaGluLeuThrAspAlaGlnIleValGlnGlnGlnLeuSerAlaProVal 320
 Db 901 GGTCTTGGGGCAGAACTTAACAGATGCTCAGATGTTGAGAGGAAAGATCTGCCCGGCT 960
 QY 321 MetValAlaIleIleProAlaArgSerTrrMetMetIleArgThrCysAlaProLysVal 340

Db 961 ATGTGTGGATCATTCGGCGCATTCGATGATGATGATTCGACTTGTGCGCTTACGTC 1020
 QY 341 PheGlyGluIleLeuGlyLysValGlyAlaGlyValThrTrrPArgLysLeuArgAspHis 360
 Db 1021 TTCGGCGAGATCTCTCGAATCTGTGAACTGGCGGACTTGGGAAATCTTCGTGATCAC 1080
 QY 361 LeuHisTrrArgLeuProAlaAlaAspValArgProLieserValArgLys 379
 Db 1081 TTGCAATTCATTCATTCGCCGACCGGATGTGCGGCCCCGATTAATTCGTTCCGAAAG 1137
 RESURF 2
 US-10-007-452-1
 ; Sequence 1, Application US/10007452
 ; Publication No. US20030093701A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Tomb, Jean-Francois
 ; APPLICANT: Bramucci, Michael G.
 ; APPLICANT: Chen, Qiong
 ; APPLICANT: Kostichka, Kristy N.
 ; TITLE OF INVENTION: Rhodococcus Cloning and Expression Vectors
 ; FILE REFERENCE: CL1709 US NA
 ; CURRENT APPLICATION NUMBER: US/10/007,452
 ; PRIOR FILING DATE: 2001-11-08
 ; PRIOR APPLICATION NUMBER: 60/254,868
 ; PRIOR FILING DATE: 2000-12-12
 ; NUMBER OF SEQ ID NOS: 30
 ; SOFTWARE: Microsoft Office 97
 ; SEQ ID NO 1
 ; LENGTH: 1140
 ; TYPE: DNA
 ; ORGANISM: Rhodococcus AM12
 ; US-10-007-452-1
 Alignment Scores:
 Pred. No.: 1,08e-213 Length: 1140
 Score: 2005.00 Matches: 379
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 14 Gaps: 0
 US-10-007-527A-2 (1-379) x US-10-007-452-1 (1-1140)
 QY 1 MetThrSerValSerAlaGluHisLeuSerGlyLysAspArgProProValLeuValSer 20
 Db 1 ATGACCAGCGTAAAGTCTTAACACCTTCCGGCAAGACCGGCTCCCTCGTGTG 60
 QY 21 SerAspLysArgGlyTLeaRgHisLeuEuArgProLysLeuGlnGlnLeThrThrSer 40
 Db 61 TCCGATTAAGCGCGCATCCGGACAGAACCTGCCAACCAATTCAACAATCACCAGTCA 120
 QY 41 GluThrPheAsnAlaCysGlyArgProLieserGlyValAsnGlyValThrIleValAsn 60
 Db 121 GAAACATTTAAAGCCCTGTGCGCGCCGATTTCTGGCGTAACGGGTGACCATTTGTCAAC 180
 QY 61 GlyProLysGlySerGlyPheGlyGlyLeuArgSerCysGlyLysGlyTrrPileCysPro 80
 Db 181 GGTTCGAAAGTTCCTGAGATTCGAGAGCCCTTCGTTCTGCGGAAAGGCTGATCTGCCCC 240
 QY 81 CysCysAlaGlyLysValGlyAlaHisArgAlaAspGluLieserGlnValAlaHis 100
 Db 241 TCGTGTGCGGAAAAAGTCGGTCACATCGTGACAGAAATTTCTCAAGTTGTGCTCAT 300
 QY 101 GlnLeuGlyThrGlySerValAlaMetValThrMetMetArgHisThrAlaGlyGln 120
 Db 301 CAACCTGGAGACTGGATCTGTGCGATGTGACGATGACCATCGCCATACAGCTGATCAG 360
 QY 121 ArgLeuHisAspLeuThrPheGlyLeuSerAlaAlaTrrPylsAlaAlaThrAsnGlyArg 140
 Db 361 CGGCTCCAGCACTATGACCTGACCTTTCGGCAGCCTGAAACCTGCCACCAACGCTCT 420
 QY 141 ArgTrrPArgThrGluArgGluMetTyrGlyCysAspGlyTyrValArgAlaValGluIle 160

```

Db      421 CGTTGGCGTACGGAAGTGAATGTAACGCTGGACGGAACACTGCCGCTGTTGAATC 480
QY      161 ThrHISGlyLYsaSngLYTrpHISValHISValHISaleuLeuMetPheSerGlyAsp 180
Db      481 ACTCAGCGAAAAAAGCGCTGGCAGTCCACGTTCAACGCGCTACTCATGTTCAATGGTGAC 540
QY      181 ValSerGluAsnIleLeuGluSerPheSerAspAlaMetCheSparGTrpThrSerLYs 200
Db      541 GTAGTGAAGAACATCTCCGATCTCTCCGATGCGATGTCATGCGTGAAGCTTCGCAAA 600
QY      201 LeuValSerLeuGlyPheAlaAlaProLeuArgAsnSerGlyLYLeuAspValArgLYs 220
Db      601 CTCGTATCTCTGGGATTTGCTGGCCACACACGTAATTCGGTGTCTCGATGTACAAAG 660
QY      221 IIEGlyGlyGluAlaAspGluValLeuAlaAlaTYrLeuThrLYsIleAsSerGlyVal 240
Db      661 ATCGCGGTGAAGCTGATCAAGTTCTCGCTGCGTATCTGACGAAATTCATGCTGGCGTT 720
QY      241 GlyMetGluValGlySerGlyAspGlyLYsSerGlyArgHISGlyAsnArgAlaProTrp 260
Db      721 GGTATGAGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 780
QY      261 GluIleAlaValAspAlaValGlyLYAspProGluAlaLeuGluLeuTrpArgGlyLys 280
Db      781 GAATTCGCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 840
QY      281 GluPheGlySerMetGlyArgArgAlaIleAlaTrpSerArgGlyLeuArgAlaArgAla 300
Db      841 GAGTTGGTTTCATGAGTGGAGTGGCGCATGCGCTGTGTGTGTGTGTGTGTGTGTGTGT 900
QY      301 GlyLeuGlyValGluLeuThrAspAlaGluIleValGluGluGluGluSerAlaProVal 320
Db      901 GGTCTGGGGGCGAGAACTAACAGATGCTCAGATGCTGAGAGAGAAATCTGCCCGGTC 960
QY      321 MetValAlaIleIleProAlaArgSerTrpMetMetIleArgThrCysAlaProTYrVal 340
Db      961 ATGGTTGCAATCATTCCTGGCGCGATGCTGATGATGATGATGATGATGATGATGATG 1020
QY      341 PheGlyGluIleLeuGlyLeuValGluAlaGlyAlaThrTrpGluAsnLeuArgAspHIS 360
Db      1021 TTCGGGGAATCCTCGGACTCGTCGAAAGCTGGCGCATTTGGGAAATCTTCGTATCAC 1080
QY      361 LeuHISTYrArgLeuProAlaAlaAspValArgProProIleIleSerValArgLYs 379
Db      1081 TTGCATTATCGATTGCCCGCAGCGGATGTGGCGCCCGCATATATCGCTTGCAGAG 1137

RESULT 3
US-10-007-527a-5/c
/ Sequence 5, Application US/10007527A
/ Publication No. US20030044807A1
/ GENERAL INFORMATION:
/ APPLICANT: Tomb, Jean-Francois
/ APPLICANT: Bramucci, Michael G.
/ APPLICANT: Cheng, Qiong
/ APPLICANT: Kostichka, Kristy N.
/ TITLE OF INVENTION: Rhodococcus Cloning and Expression Vectors
/ CURRENT APPLICATION NUMBER: US/10/007,527A
/ PRIOR FILING DATE: 2001-12-05
/ PRIOR APPLICATION NUMBER: 60/254,868
/ NUMBER OF SEQ ID NOS: 30
/ SOFTWARE: Microsoft Office 97
/ SEQ ID NO 5
/ LENGTH: 6334
/ TYPE: DNA
/ ORGANISM: Rhodococcus AN12
US-10-007-527a-5

Alignment Scores:
Pred. No.: 1 02e-212 Length: 6334
Score: 2005.00 Matches: 379

```

```

Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 14 Gaps: 0
US-10-007-527a-2 (1-379) x US-10-007-527a-5 (1-6334)

QY      1 MetThrSerValSerAlaGluHISLeuSerGlyLYsAspArgProProValLeuValSer 20
Db      3051 ATGACCAAGCGTAAGTGTCTGAACACCTTTCGGCAAAAGCCGGGCTCCGCTCGTGTG 2992
QY      21 SerAspLYsArgGlyLYLeuArgHISGluLeuArgProLYsLeuGluIleThrThrSer 40
Db      2991 TCCGATTAACCGCGGATCCGAGACAACTCCGACCCCAACTTCAACAAATCAACAGTCA 2932
QY      41 GluThrPheAsnAlaCysGlyLYsArgProIleSerGlyValAsnGlyValThrIleValAsn 60
Db      2931 GAAACATTAAAGCTGTGTGGCCGGCCGATTTTGGCGGTAAACGGGTGTACATGTGTCAAC 2872
QY      61 GlyProLYsGlySerGlyPheGlyGlyLeuArgSerCysGlyLYsGlyTYrIleCysPro 80
Db      2871 GGTCCGAAAGTTCTGATTCGAGGCTTCGTTCTGTGTGTGTGTGTGTGTGTGTGTGTGT 2812
QY      81 CysCysAlaGlyLYsValGlyValAlaHISArgAlaAspGluIleSerGluValValAlaHIS 100
Db      2811 TCTGTGTGGGAAAAAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2752
QY      101 GluLeuGlyTYrGlySerValAlaMetValThrMetThrMetArgHISThrAlaGlyGln 120
Db      2751 CAATCGGAGCTGTGATCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2692
QY      121 ArgLeuHISAspLeuTrpThrGlyLeuSerAlaAlaTrpLYsAlaAlaThrAsnGlyArg 140
Db      2691 CGGCTCCAGACCTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2632
QY      141 ArgTrpArgThrGluArgGluMetTYrGlyCysAspGlyTYrValArgAlaValGluIle 160
Db      2631 CGTTGGCGTACGGAACGTGAATGTACCGCTCGACGAGTACGTGGCGCTGTGTGAATC 2572
QY      161 ThrHISGlyLYsAsnGlyTYrHISValHISValHISaleuLeuMetPheSerGlyAsp 180
Db      2571 ACTCAGCGAAAAAAGCGCTGGCAGCTTCACGTTACGCGCTCAATGTTCAATGTGTGTAC 2512
QY      181 ValSerGluAsnIleLeuGluSerPheSerAspAlaMetPheAspArgTrpThrSerLYs 200
Db      2511 GGAATGGAACATCTCGAATCTCTCTCGATGATGATGATGATGATGATGATGATGATG 2452
QY      201 LeuValSerLeuGlyPheAlaAlaProLeuArgAsnSerGlyLYLeuAspValArgLYs 220
Db      2451 CTCGTATCTCTGGGATTTGCTGGCCACACGTAATTCGGTGTGTGTGTGTGTGTGTGTGT 2392
QY      221 IIEGlyGlyGluAlaAspGluValLeuAlaAlaTYrLeuThrLYsIleAsSerGlyVal 240
Db      2391 ATCGGCGGTGAAGCTGATCAAGTTCTCGTGGTATCTGACGAAATTCATCTGCGCTT 2332
QY      241 GlyMetGluValGlySerGlyAspGlyLYsSerGlyArgHISGlyAsnArgAlaProTrp 260
Db      2331 GGTATGAGGTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2272
QY      261 GluIleAlaValAspAlaValGlyLYsAspProGluAlaLeuGluLeuTrpArgGlyLys 280
Db      2271 GAATTCGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2212
QY      281 GluPheGlySerMetGlyArgArgAlaIleAlaTrpSerArgGlyLeuArgAlaArgAla 300
Db      2211 GAGTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2152
QY      301 GlyLeuGlyValGluLeuThrAspAlaGluIleValGluGluGluGluSerAlaProVal 320
Db      2151 GGTCTTGGGCGAAGTAAACATGCTCAGATGCTTGAACAGAAAGAAATCTCCCGGTC 2092
QY      321 MetValAlaIleIleProAlaArgSerTrpMetMetIleArgThrCysAlaProTYrVal 340

```

```
Db      2091 ATGGTTGCATATTCGGCGCGGCGATGATGATGATTCGGATCTTGCGCCTTAACGTC 2032
QY      341 PhegiyguilleleuengllyeuvalaglyalathrtpgluasnuargaspHis 360
Db      2031 TTCGGGAGATCCTCGGATCTCGATCGAGCTGGGGGAGACTTGGGAAAATCTTGATGATCAC 1972
QY      361 LeuHsiTYrAArgleuProAlaAlaaspValArgProProllelleserValArglys 379
Db      1971 TTGCATATTCGATTCGCCCGAGCGGATGTGCGGCGCCCGCATATATCGGTTGCGCAAG 1915

RESULT 4
US-10-007-452-5/c
; Sequence 5, Application US/10007452
; Publication No. US20030093701A1
; GENERAL INFORMATION:
; APPLICANT: Tomb, Jean-Francois
; APPLICANT: Bramucci, Michael G.
; APPLICANT: Cheng, Qiong
; APPLICANT: Kostichka, Kristy N.
; TITLE OF INVENTION: Rhodococcus Cloning and Expression Vectors
; FILE REFERENCE: CL1709 US NA
; CURRENT APPLICATION NUMBER: US/10/007,452
; PRIOR FILING DATE: 2001-11-08
; PRIOR APPLICATION NUMBER: 60/254,868
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 5
; LENGTH: 6334
; TYPE: DNA
; ORGANISM: Rhodococcus AN12
US-10-007-452-5

Alignment Scores:
Pred. No.:      1,02e-212      Length:      6334
Score:          2005.00      Matches:      379
Percent Similarity: 100.00%      Conservative: 0
Best Local Similarity: 100.00%      Mismatches: 0
Query Match:     100.00%      Indels:      0
DB:              14          Gaps:          0

US-10-007-527a-2 (1-379) x US-10-007-452-5 (1-6334)
QY      1 MetThrSerValSerAlaGluHisleuSerGlyValAspArgProProValleuValSer 20
Db      3051 ATGACACGAGCTAAGTCTGTAACACCTTCCGGCAAGACCGGCTCCCGCTCGATGCG 2992
QY      21 SerAspLysArgGlyIleArgHisGluLeuArgProLysLeuGlnGlnIleThrThrSer 40
Db      2991 TCCGATTAACCGCGGATCCGGGACGAACTGCGACCCCAACTTCACAAATTCACACGCTCA 2932
QY      41 GluThrPheAsnAlaCysGlyArgProIleSerGlyValAsnGlyValThrIleValAsn 60
Db      2931 GAAACATTTAACCGCTGTGGCCGCGGCGGATTTCTGGCGTGAACGCTGTGACATTGTCAAC 2872
QY      61 GlyProLysGlySerGlyPheGlyGlyLeuArgSerCysGlyValysGlyTyrPileCysPro 80
Db      2871 GGTCCGAAAGGTTCTGATTCGAGAGGCTTCGTTCTCGCGAAAGGGCTGATCTGCCCC 2812
QY      81 CysCysAlaGlyIlyValAlaGlyAlaHisArgAlaAspGluIleSerGlnValValaHis 100
Db      2811 TGTGTGCGGGAAGAAATCGGTGCACATCGGACGAGCAAAATTTCTCAAGTTGTGTCTCAT 2752
QY      101 GlnLeuGlyThrGlySerValAlaMetValThrMetThrMetArgHisThrAlaGlyGln 120
Db      2751 CAACCTGGGAGCTGATCTGTGGCGATGTGACATACATACATGCGCATACAGCTGTGACG 2692
QY      121 ArgLeuHisAspLeuTyrThrGlyLeuSerValAlaIleTyrLysAlaIleThrAsnGlyArg 140
Db      2691 CGGCTCCACGACCTATGACATGACATTTCCGACGCTGGAAGAACTGCGACCAACGAGTCGT 2632
QY      141 ArgTTrpArgThrGluArgGluMetTyrGlyCysAsnGlyTyrValArgAlaValGluIle 160
```

```
Db      2631 CGTTGGCGGTACGGAACGTAATGTACAGGCTTGCCACGGAATACGTCGCGCTTGGAATC 2572
QY      161 ThrHisGlyLysAsnGlyTyrPheValHisValHisAlaLeuLeuMetPheSerGlyAsp 180
Db      2571 ACTCACGGAAAAAAGCGGTGGCAAGTCCACGTTACCGGGCTACTCATGTTCAAGTGTAC 2512
QY      181 ValSerGluAsnIleLeuGluSerPheSerAspAlaMetPheAspArgTyrThrSerLys 200
Db      2511 GTGAGTGAACATCTCCGATTCCTTCGCGATGCGATGTTGATCGATCGATGCACTTCCAA 2452
QY      201 LeuValSerLeuGlyPheAlaAlaProleuArgAsnSerGlyValLeuAspValArglys 220
Db      2451 CTCGATCTTCGGATTTGCTGCGCCACTTACGTAATTCGGATGATCTCGATGTACGAAG 2392
QY      221 IleGlyGlyGluAlaAspGlnValleuAlaIleTyrLeuThrLysIleAlaSerGlyVal 240
Db      2391 ATCCGCGGTGAAGCTGATCAAGTTCTGCTGATGATCGAGAAATTCGATCTGCGCTT 2332
QY      241 GlyMetGluValGlySerGlyAspGlyLysSerGlyArgHisGlyAsnArgAlaProTyr 260
Db      2331 GGTATGAGAGTTGGTGTGTGGCGACGGAAGAAAGTGTGCACATGCGCACCCGTCG 2272
QY      261 GluIleAlaValAlaAspAlaValGlyValAspProGlnAlaLeuGluLeuTyrArgGluPhe 280
Db      2271 GAATTCGCTGTGATGACGTGGCGGGATTCACAAAGCTTGGAACTGTGGCGAATTT 2212
QY      281 GluPheGlySerMetGlyArgArgAlaIleAlaTyrSerArgGlyLeuArgAlaArgAla 300
Db      2211 GATTTGGTTCGATGTGACGTGGGCAATCCGATGTCGCGGATGTCGTCGTCGCGAGCT 2152
QY      301 GlyLeuGlyAlaGluLeuThrAspAlaGlnIleValGluGlnGluSerAlaProVal 320
Db      2151 GGTCTTGGCGAGAACTACAGATGCTCAGATCGTTGAGCAGGAAGATCTGCCCGGTC 2092
QY      321 MetValAlaIleIleProAlaArgSerTrpMetMetIleArgThrCysAlaProTyrVal 340
Db      2091 ATGGTTGGATCATTCGCGCGGATCGAGATGATGATTCGACTTGTGCGCCTTAACGTC 2032
QY      341 PhegiyguilleleuengllyeuvalaglyalathrtpgluasnuargaspHis 360
Db      2031 TTCGGGAGATCCTCGGATCTCGATCGAGCTGGGGGAGACTTGGGAAAATCTTGATGATCAC 1972

RESULT 5
US-10-007-527a-7/c
; Sequence 7, Application US/10007527a
; Publication No. US20030044807A1
; GENERAL INFORMATION:
; APPLICANT: Tomb, Jean-Francois
; APPLICANT: Bramucci, Michael G.
; APPLICANT: Cheng, Qiong
; APPLICANT: Kostichka, Kristy N.
; TITLE OF INVENTION: Rhodococcus Cloning and Expression Vectors
; FILE REFERENCE: CL1709 US NA
; CURRENT APPLICATION NUMBER: US/10/007,527a
; PRIOR FILING DATE: 2001-12-05
; PRIOR APPLICATION NUMBER: 60/254,868
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 7
; LENGTH: 9652
; TYPE: DNA
; ORGANISM: Plasmid pRHB17
US-10-007-527a-7

Alignment Scores:
Pred. No.:      1.76e-212      Length:      9652
Score:          2005.00      Matches:      379
Percent Similarity: 100.00%      Conservative: 0
```

Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 14 Gaps: 0

US-10-007-527a-2 (1-379) x US-10-007-527a-7 (1-9652)

```
QY      1 MetThSerValSerAlaGluHisLeuSerGlyLysAspArgProProValLeuValSer 20
DB      5052 ATGACCGAGGTGAAGTGTGAACACCTTTCGGCAAAAGCGGCTCCGCTCCGTGTGCG 4993

QY      21 SerAspLysArgGlyIleArgHisGluLeuArgProLysLeuGlnGlnIleThrThrSer 40
DB      4992 TCCGATTAAGCGCGGCAATCCGGCAGCAATCGCAACCAATTCTCAACAAATCACCGTCA 4933

QY      41 GluThrPheAsnAlaCysGlyArgProIleSerGlyValAsnGlyValThrIleValAsn 60
DB      4932 GAAACATTTTAACGCTGTGCGCGGCAATTTCTGCGGTGAACGGTGTGACCATGTGTCAAC 4873

QY      61 GlyProLysGlySerGlyPheGlyGlyLeuArgSerGlyLysGlyTTP11eCysPro 80
DB      4872 GGTCCGAAAGGTTCTGGATTGCGAGGCTTCGTTCCGCGAAAGGGCTGATCTGCCCC 4813

QY      81 CysCysAlaGlyLysValGlyAlaHisArgAlaAspGluIleSerGlnValAlaHis 100
DB      4812 TGTCTGCGCGGAAAGTCCGTGCACATCGCAGACGAAATTTCTCAAGTTGTGCTCAT 4753

QY      101 GlnLeuGlyThrGlySerValAlaMetValThrMetThrValArgHisGlyValGln 120
DB      4752 CAACCTCGGAGCTGATCTGTGCGATGTGACGATGACATGCGCCATACAGCTGTGTGACG 4693

QY      121 ArgLeuHisAspLeuThrPheThrGlyLeuSerAlaAlaThrIleValAlaHisArg 140
DB      4692 CGGCTCCACGACCTATGAGCTGAGACTTTCGCGAGCCCTCGAAAGCTGTGCAACCGGTCTG 4633

QY      141 ArgThrArgThrGlnArgGluMetTyrGlyCysAspGlyTyrValArgAlaValGluIle 160
DB      4632 CGTTGGCGTACGGAACGTGAATGTACGCTGCGACGAGATACGTGCGCGCTGTGGAATTC 4573

QY      161 ThrHisGlyLysAsnGlyTTPH1sValHisValHisAlaLeuLeuMetPheSerGlyAsp 180
DB      4572 ACTCACGGAAGAAACCGCTGCGACGTCCACGTTACGCGCTACGTCATGTCAGTGGTGCAC 4513

QY      181 ValSerGluAsnGlnIleLeuGluSerPheSerAspAlaMetPheAspArgTTPH1SerLys 200
DB      4512 GTAGGTGAAGAACATCTCCGATCTCTTCGATGCGATGCTTCATCTGCTGAGCTTCCAA 4453

QY      201 LeuValSerLeuGlyPheAlaAlaProLeuArgAsnSerGlyGlyLeuAspValArgLys 220
DB      4452 CTGATCTCTGGGATTTTCTGCGGCCCTACGTAATTCGGGTGTCTCGATGTACGAAAG 4393

QY      221 IleGlyGlyValAlaAspGlnValLeuAlaAlaTyrIleuThrLysIleAlaSerGlyVal 240
DB      4392 ATCGGCGGTGAAGCTGATCAAGTTCTCGCTCGGATCTGACGAAATTCATCTGGGCTT 4333

QY      241 GlyMetGluValGlySerGlyLysArgGlyLysSerGlyArgHisGlyValAsnAlaGlyProTyr 260
DB      4332 GGATATGAAGGTGTGTATGTGCGACGCGGAAAGTGTGTGACATGGCAACCTGTGACCTCTGG 4273

QY      261 GluIleAlaValAspAlaValGlyLysAspProGlnAlaLeuGlnLeuTTPArgGluPhe 280
DB      4272 GAAATCCCTGTTGATGCAGTGGGCGGGAATCCACAAACGTTGGACTGTGTGCGAAGATTTT 4213

QY      281 GluPheGlySerMetGlyArgArgAlaIleAlaTTPSerArgGlyLeuArgAlaArgAla 300
DB      4212 GAGTTTGTTTCGATGGGACGTCGGGCAATCGCGTGTGCTCGGTGATTTGGGTGCCGAGCT 4153

QY      301 GlyLeuGlyValaGluLeuThrAspAlaGlnIleValGlnGlnGlnGlnGlnGlnGlnGln 320
DB      4152 GGTCTTGCGGCAACAATCAAGATGCTCAAGATGTTGACGAGAAAGATCTGCCCCGCTC 4093

QY      321 MetValAlaIleIleProAlaArgSerTTPMetMetIleArgThrCysAlaIleProTyrVal 340
DB      4092 ATGTGTGCGATCATCTCGGCGGACGATGATGTATGTATGTGACTGTGTGCGCTTACGTC 4033
```

```
QY      341 PheGlyGlnIleLeuGlyLeuValGlyAlaGlyAlaThrTTPGluAsnLeuArgAspHis 360
DB      4032 TTCGGAGAGATCTCTGAGCTGCGAAGCTGGCGCACTTGGGAAATCTTCGTGATCTAC 3973

QY      361 LeuHisTyrArgLeuProAlaAlaAspValArgProProIleIleSerValArgLys 379
DB      3972 TTGCATTAATCATGTGCGCGGAGGATGTGCGGCCCCCGAATATATGATGTGCGACAG 3916
```

RESULT 6

```
US-10-007-452-7/c
; Sequence 7, Application US/10007452
; Publication No. US20030093701A1
; GENERAL INFORMATION:
; APPLICANT: Tomb, Jean-Francois
; APPLICANT: Bramucci, Michael G.
; APPLICANT: Cheng, Olong
; APPLICANT: Kostichka, Kristy N.
; TITLE OF INVENTION: Rhodococcus Cloning and Expression Vectors
; FILE REFERENCES: CL1709 US NA
; CURRENT APPLICATION NUMBER: US/10/007,452
; PRIOR FILING DATE: 2001-11-08
; PRIOR APPLICATION NUMBER: 60/254,868
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 7
; LENGTH: 9652
; TYPE: DNA
; ORGANISM: Plasmid pRHB17
US-10-007-452-7
```

Alignment Scores:

```
Pred. No.: 1,766-212 Length: 9652
Score: 2005.00 Matches: 379
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 14 Gaps: 0
```

US-10-007-527a-2 (1-379) x US-10-007-452-7 (1-9652)

```
QY      1 MetThSerValSerAlaGluHisLeuSerGlyLysAspArgProProValLeuValSer 20
DB      5052 ATGACCGAGGTGAAGTGTGAACACCTTTCGGCAAAAGCGGCTCCGCTCCGTGTGCG 4993

QY      21 SerAspLysArgGlyIleArgHisGluLeuArgProLysLeuGlnGlnIleThrThrSer 40
DB      4992 TCCGATTAAGCGCGGCAATCCGGCAGCAATCGCAACCAATTCTCAACAAATCACCGTCA 4933

QY      41 GluThrPheAsnAlaCysGlyArgProIleSerGlyValAsnGlyValThrIleValAsn 60
DB      4932 GAAACATTTTAACGCTGTGCGCGGCAATTTCTGCGGTGAACGGTGTGACCATGTGTCAAC 4873

QY      61 GlyProLysGlySerGlyPheGlyGlyLeuArgSerGlyLysGlyTTP11eCysPro 80
DB      4872 GGTCCGAAAGGTTCTGGATTGCGAGGCTTCGTTCCGCGAAAGGGCTGATCTGCCCC 4813

QY      81 CysCysAlaGlyLysValGlyAlaHisArgAlaAspGluIleSerGlnValAlaHis 100
DB      4812 TGTCTGCGCGGAAAGTCCGTGCACATCGCAGACGAAATTTCTCAAGTTGTGCTCAT 4753

QY      101 GlnLeuGlyThrGlySerValAlaMetValThrMetThrValArgHisGlyValGln 120
DB      4752 CAACCTCGGAGCTGATCTGTGCGATGTGACGATGACATGCGCCATACAGCTGTGTGACG 4693

QY      121 ArgLeuHisAspLeuThrPheThrGlyLeuSerAlaAlaThrIleValAlaHisArg 140
DB      4692 CGGCTCCACGACCTATGAGCTGAGACTTTCGCGAGCCCTCGAAAGCTGTGCAACCGGTCTG 4633

QY      141 ArgThrArgThrGlnArgGluMetTyrGlyCysAspGlyTyrValArgAlaValGluIle 160
DB      4632 CGTTGGCGTACGGAACGTGAATGTACGCTGCGACGAGATACGTGTGCGCTGTGGAATTC 4573
```


QY 341 PheGlyValIleuenglyLeuValGluAlaGlyValAlaThrTrpGluSubLeuArgAspHis 360
Db 5621 TTGGGCGAGATCCTCGGACTCGTGAAGCTGGGCGCACTTGGGAAAATCTTGGATGAC 5562
QY 361 LeuHisIleYrArgLeuProAlaAlaAspValArgProProlIleIleSerValArgLys 379
Db 5561 TTGCATTATCGATTGGCCCGCAGCGGAGTGTGGGCCCCCGATTAATATCGTTCCGAG 5505

RESULT 8
US-10-007-452-6/c
; Sequence 6, Application US/10007452
; Publication No. US20030093701A1
; GENERAL INFORMATION:
; APPLICANT: Tomb, Jean-Francois
; APPLICANT: Bramucci, Michael G.
; APPLICANT: Cheng, Qiong
; APPLICANT: Kostichka, Kristy N.
; TITLE OF INVENTION: Rhodococcus Cloning and Expression Vectors
; FILE REFERENCES: C11709 US NA
; CURRENT APPLICATION NUMBER: US/10/007,452
; PRIOR FILING DATE: 2001-11-08
; PRIOR APPLICATION NUMBER: 60/254,868
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 6
; LENGTH: 11241
; TYPE: DNA
; ORGANISM: Plasmid PRHBR17
US-10-007-452-6

Alignment Scores:
Pred. No.: 2.15e-212 Length: 11241
Score: 2005.00 Matches: 379
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 14 Gaps: 0

US-10-007-527A-2 (1-379) x US-10-007-452-6 (1-11241)

QY 1 MetThrSerValSerIleGluHisIleuSerGlyValAspArgProProValIleuValSer 20
Db 6641 ATACCAAGCGTAAAGTGTGAACACCTTTCGGCAAAACCGGCTCCGCTCGTGTG 6582
QY 21 SerAspLysArgGlyIleArgHisGluLeuArgProLysLeuGlnIleThrThrSer 40
Db 6581 TCCGATTAAGCGCGGATCCGCGACGAATCCGACCCAACTTCAACAAATCACACGTC 6522
QY 41 GluThrPheAsnAlaCysGlyArgProIleSerGlyValAsnGlyValThrIleValAsn 60
Db 6521 GAAACATTTAAACCGCTGTGGCCGCGCGATTTCTGGCGTGAACGATGTGACCATTTGTC 6462
QY 61 GlyProLysGlySerGlyPheGlyGlyLeuArgSerGlyLysGlyTTPPleCysPro 80
Db 6461 GGTCCGAAAGGTTCTGATTCGAGGCGCTTCGCTCCGCGAAAGGCGGATCTGCCCC 6402
QY 81 CysCysAlaGlyLysValGlyAlaHisArgAlaAspGluIleSerGlnValAlaHis 100
Db 6401 TGCCTGCCGGAAGAAAGTCGATCGATCGATCGACGACGAATTTCTCAAGTTGTTCAT 6342
QY 101 GlnLeuGlyThrGlySerValAlaMetValThrMetThrMetArgHisThrAlaGlyGln 120
Db 6341 CAACCTCGGACTGGATCTGTTCGATGTGATGATGATGATGATGATGATGATGATGATG 6282
QY 121 ArgLeuHisAspLeuThrTrpGlyLeuSerAlaAlaTrpLysAlaAlaThrAsnGlyArg 140
Db 6281 CGGCTCCACGACCTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 6222
QY 141 ArgTrpArgThrGluArgGlyMetGlyCysAspGlyTyrValArgAlaValGluIle 160
Db 6221 CGTTGGCGTACGAGAACGTGAATGTACGCTGCGACGATACGTGCGCGCTGTGAATC 6162

QY 161 ThrHisGlyLysAsnGlyTTPPleValHisValHisAlaLeuLeuMetPheSerGlyAsp 180
Db 6161 ACTCAAGGAAAAAAGCGCTGGCAAGTCCAGCTTCCAGCCCTACTCATGTTCAAGTGTGAC 6102
QY 181 ValSerGluAsnIleuenglyLeuSerPheSerAspAlaMetPheAspArgTrpThrSerLys 200
Db 6101 GTGAGGAGAAACATCTCCAAATCTTCTCGATGCGATTTGATGCGGACTTCCAA 6042
QY 201 LeuValSerLeuGlyPheAlaAlaProLeuArgAsnSerGlyLysAspValArgLys 220
Db 6041 CTCGATCTCTGGATTTGCTGGCCACATACGTAATTCGGGTGTCTCGATGTACGAAG 5982
QY 221 IleGlyGlyValAlaAspGlnValLeuAlaAlaTyrLeuThrLysIleAlaSerGlyVal 240
Db 5981 ATCGGGGGAAGCTGATCAAGTTCGCTGCTATCTGACGAAAATTCGATCTGGCGGT 5922
QY 241 GlyMetGluValGlySerGlyAspGlyLysSerGlyArgHisGlyAsnArgAlaProTrp 260
Db 5921 GGTATGAGGTTGT 5862
QY 261 GluIleAlaValAspAlaValGlyGlyAspProGlnAlaLeuGluLeuTrpArgGluPhe 280
Db 5861 GAAATCGCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 5802
QY 281 GluPheGlySerMetGlyArgArgAlaIleAlaTTPSerArgGlyLeuArgAlaArgAla 300
Db 5801 GAGTTGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 5742
QY 301 GlyLeuGlyAlaGluLeuThrAspAlaGlnIleValGluGlnGluSerAlaProVal 320
Db 5741 GGTCTGGGCGAAGATTAAACAGATGCTCGATGCTTGAAGAGAAAGATTTGCCCGGTC 5682
QY 321 MetValAlaIleIleProAlaArgSerTrpMetMetIleArgThrCysAlaProVal 340
Db 5681 ATGTTGTGATCATTTCCGCGCGATGCTGTGATGATGATGATGATGATGATGATGATGAT 5622
QY 341 PheGlyGluIleLeuGlyLeuValGluAlaGlyAlaThrTrpGluSubLeuArgAspHis 360
Db 5621 TTCCGGGAGATCCTCGGACTCGTCCGAGCTGGCGCACTTGGGAAAATCTTCCGTGATC 5562

QY 361 LeuHisIleYrArgLeuProAlaAlaAspValArgProProlIleIleSerValArgLys 379
Db 5561 TTGCATTATCGATTGGCCCGCAGCGGAGTGTGGGCCCCCGATTAATATCGTTCCGAG 5505

RESULT 9
US-09-835-381-5
; Sequence 5, Application US/09835381
; Patent No. US20020045223A1
; GENERAL INFORMATION:
; APPLICANT: SUGA, Mikiko
; APPLICANT: ASAKURA, Yoko
; APPLICANT: MORI, Yukiko
; APPLICANT: ITO, Hisao
; APPLICANT: KURAHASHI, Osamu
; TITLE OF INVENTION: ARGININE REPRESSOR DEFICIENT STRAIN OF CORYNEFORM BACTERIUM
; FILE REFERENCE: 206018US0
; CURRENT APPLICATION NUMBER: US/09/835,381
; PRIOR FILING DATE: 2001-04-17
; PRIOR APPLICATION NUMBER: JP2000-129167
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 5
; LENGTH: 4447
; TYPE: DNA
; ORGANISM: Brevibacterium lactofermentum
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1318)..(2598)
; OTHER INFORMATION:
US-09-835-381-5

Alignment Scores:

Pred. No.:	6,05e-26	Length:	4447
Percent Similarity:	326.00	Matches:	101
Best Local Similarity:	47.75%	Conservative:	58
Query Match:	30.33%	Mismatches:	133
	16.26%	Indels:	42
		Gaps:	14

US-10-007-527A-2 (1-379) x US-09-835-381-5 (1-4447)

```
QY      34  LeuGlnGlnIleThrSerGluThrPheAsnAlaCysGlyArgProIleSerGlyVal 53
      1396  ATGTACAGATCCACCAATAGTAAAGCTCGCGCGGCGCATGAGTG-GGCGAGGACGCA 1454
      54  Asn-----GlyValThrIleVal-AsnGlyProIlySerGlyPheGlyLeuArg 71
      1455  AGCTGTTCGGGTCCGTGGTCTAACGGT-----GCTTCGACATTGAGGGTCTGCA 1508
      71  gSerCysGlyLysGlyTrrIleCysProCysSalaglyLysValGlyValahisArgAl 91
      1509  AAACCTCACTCGCGCGGGGCGACCTCGGCTGAATTGAAAGTCATGGCGGACGCCG 1568
      91  aAspGluIleSerGlnValValahisGlnLeuGlyThrGlySerValAlaMetValTh 111
      1569  CATTAAGCTGCTATTGCTACTAAGATCACTTGGCGCGGCGGCGCTCATGATGTT 1628
      111  rMet---ThrMetArgHisThrAlaGlyGlnArgLeuHisAspLeuTrrThrGlyLeuSe 130
      1629  TGTGGGCACTGTTGCACACACCGCTCAAGTCATTTGCCAGAGTTGAAGCGGATTTAA 1688
      130  rAlaAlaTrrPlySalAlaAlaThrAsnGlyArgArgTrrPArgThrGluArgGluMetYrgI 150
      1689  GACTGGCTACTCTTCGATGCTGAAACATCTCACTGGTGAAGAAAGACGTGACGCGTA 1748
      150  yCysAspGlyTrrValArgAlaValGluIleThrHisGly---LysAsnGlyTrrPhisVa 169
      1749  GGTGGAGCACACCTATAGTACTATGAGTCAAGTCAACAGACTCTTGGCGGACGCTTG 1808
      169  lHisValHisAlaLeuLeuMetPheSerGlyAspValSerGluAsnIleLeuGlnSerPh 189
      1809  GACCCGCAACATGCTGTGTTCTTGATCGCTCACTTCCTGCACATCAACTCAAGCGG 1868
      189  eSerAspAlaMetPheAspArgTrrPrrSerIlySleValSerLeuGlyPheAlaAlaPr 209
      1869  TAGAGATTCATGTTTCCCGCTGCTGCTGCGTGAAGCGGCTATGACGCCG 1928
      209  oLeuArgAsnSerGly---GlyLeuAsp---ValArgLysIleGlyGlyGluAlaAspG 227
      1929  ACTGCGTGAACACGCGGCTCAAACTTGATCAGGTGTCTACCTGGGCTGAGACGCTG 1988
      227  nValLeuAlaAlaIleThrLeuThrIlySleAlaSerGlyValGlyMetGlyValGlySerG 247
      1989  A---ATGGCAACTTACTCCGTG---GGCATGTCTCAGGAAGTCACTGAGTGCCTC 2036
      247  yAspGlyLysSerGlyArgHisGlyAsnArgAlaProTrrPgluIleAlaValAspAlaVa 267
      2037  CGCTACATAAAACCGCGCTAAGCGGTGCTACACCGCTTCAGATG---TTGATATGTT 2093
      267  lGlyGlyAspProGlnAlaLeuGlu-----LeuTrrPArgI 279
      2094  GGCGCATCAAAAGCACCGCGCGGAGATATGACGCTGTTTGGTGGCTGCGTGGCGCTGA 2153
      279  uPheGluPheGlySerMetGlyArgArgAlaIleAlaIleThrSerArgIlyLeuArgAlaAr 299
      2154  GTATGAGGTGTGTTCTAAAACTCGCGTTG---TCTGTGTCACTGAGGG-----GCTTA 2204
      299  gAlaGlyLeuGlyValaGluLeuThrAspAlaGlnIle-----ValGluGlnGlu-- 315
      2205  GCGGTCTTTGGGCAATTGATTACATAGACGCTGATGTACGTGTAAGTAAGAAAGAACT 2264
      316  -----GluSerAlaProValMetValAl 323
      2265  GTACAAAGCTCGCGGCTGTGAAAGACGCGGAGCGGATGAAATCAACCGCGTGTGCTGTGC 2324
```

```
QY      323  alleIleProAlaArgSerTrrPrrMetIleArgThr 335
      2325  TTTGGTGAAGCCGATGATTTGGAACATGATTAGCT 2361
```

RESULT 10

```
US-10-196-232-5
; Sequence 5, Application US/10196232
; Publication No. US20030113899A1
; GENERAL INFORMATION:
; APPLICANT: YAMAGUCHI, MIKIKO
; APPLICANT: ITO, HISAO
; APPLICANT: GUNOI, YOSHITA
; APPLICANT: YASUEDA, HISASHI
; TITLE OF INVENTION: METHOD FOR PRODUCING L-ARGININE
; FILE REFERENCE: 225391USO
; CURRENT APPLICATION NUMBER: US/10/196,232
; PRIORITY FILING DATE: 2002-07-17
; PRIOR APPLICATION NUMBER: JP 2001-224586
; NUMBER OF SEQ ID NOS: 35
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 5
; LENGTH: 4447
; TYPE: DNA
; ORGANISM: Brevibacterium lactofermentum
US-10-196-232-5
```

Alignment Scores:

Pred. No.:	6,05e-26	Length:	4447
Percent Similarity:	326.00	Matches:	101
Best Local Similarity:	47.75%	Conservative:	58
Query Match:	30.33%	Mismatches:	133
	16.26%	Indels:	42
		Gaps:	14

US-10-007-527A-2 (1-379) x US-10-196-232-5 (1-4447)

```
QY      34  LeuGlnGlnIleThrSerGluThrPheAsnAlaCysGlyArgProIleSerGlyVal 53
      1396  ATGTACAGATCCACCAATAGTAAAGCTCGCGGGTGCATGAGTG-GGCGAGGACGCA 1454
      54  Asn-----GlyValThrIleVal-AsnGlyProIlySerGlyPheGlyLeuArg 71
      1455  AGCTGTTCGGGTCCGTGGTCTAACGGT-----GCTTCGACATTGAGGGTCTGCA 1508
      71  gSerCysGlyLysGlyTrrIleCysProCysSalaglyLysValGlyValahisArgAl 91
      1509  AAACCTCACTCGCGCGGGGCGACCTCGGCTGAATTGAAAGTCATGGCGGACGCCG 1568
      91  aAspGluIleSerGlnValValahisGlnLeuGlyThrGlySerValAlaMetValTh 111
      1569  CATTAAGCTGCTATTGCTACTAAGATCACTTGGCGCGGCGGCGCTCATGATGTT 1628
      111  rMet---ThrMetArgHisThrAlaGlyGlnArgLeuHisAspLeuTrrThrGlyLeuSe 130
      1629  TGTGGGCACTGTTGCACACACCGCTCAAGTCATTTGCCAGAGTTGAAGCGGATTTAA 1688
      130  rAlaAlaTrrPlySalAlaAlaThrAsnGlyArgArgTrrPArgThrGluArgGluMetYrgI 150
      1689  GACTGGCTACTCTTCGATGCTGAAACATCTCACTGGTGAAGAAAGACGTGACGCGTA 1748
      150  yCysAspGlyTrrValArgAlaValGluIleThrHisGly---LysAsnGlyTrrPhisVa 169
      1749  GGTGGAGCACACCTATAGTACTATGAGTCAAGTCACTTGGCGGCGGCGCTCATGATG 1808
      169  lHisValHisAlaLeuLeuMetPheSerGlyAspValSerGluAsnIleLeuGlnSerPh 189
      1809  GACCCGCAACATGCTGTGTTCTTGATCGCTCACTTCCTGCACATCAACTCAAGCGG 1868
      189  eSerAspAlaMetPheAspArgTrrPrrSerIlySleValSerLeuGlyPheAlaAlaPr 209
      1869  TAGAGATTCATGTTTCCCGCTGCTGCTGCTGCTGTTAAGCGCGGATGAGACGCCG 1928
```

34 LeuGlnGlnIleThrThrSerGluThrPheAsnAlaCysGlyArgProIleSerGlyVal 53

RESULT 12
US-10-196-232-6
; Sequence 6, Application US/10196232
; Publication No. US20030113899A1
; GENERAL INFORMATION:
; APPLICANT: YAMAGUCHI, MIKIYO

```

1  APPLICANT: ITO, HISAO
2  APPLICANT: GUNJI, YOSHITKA
3  APPLICANT: YASUEDA, HISASHI
4  TITLE OF INVENTION: METHOD FOR PRODUCING L-ARGININE
5  FILE REFERENCE: 225391US0
6  CURRENT APPLICATION NUMBER: US/10/196,232
7  CURRENT FILING DATE: 2002-07-17
8  PRIOR APPLICATION NUMBER: JP 2001-224586
9  PRIOR FILING DATE: 2001-07-25
10 NUMBER OF SEQ ID NOS: 35
11 SOFTWARE: PatentIn version 3.1
12 SEQ ID NO 6
13
14 LENGTH: 4447
15
16 TYPE: DNA
17
18 ORGANISM: Brevibacterium lactofermentum
19
20 FEATURE:
21 NAME/KEY: CDS
22 LOCATION: (1318)..(2598)
23 OTHER INFORMATION:
24
25 OS-10-196-232-6

```


;/ CURRENT APPLICATION NUMBER: US/10/261,481
;/ CURRENT FILING DATE: 2002-10-02
;/ PRIOR APPLICATION NUMBER: US 09/826,191
;/ PRIOR FILING DATE: 2001-04-05
;/ PRIOR APPLICATION NUMBER: US 60/194,627
;/ PRIOR FILING DATE: 2000-04-05
;/ NUMBER OF SEQ ID NOS: 10
;/ SOFTWARE: PatentIn version 3.0
;/ SEQ ID NO 6
;/ LENGTH: 2401
;/ TYPE: DNA
;/ ORGANISM: Ketogulonigenium
US-10-261-481-6

Alignment Scores:

Pred. No.:	2,15e-21	Length:	2401
Score:	282.00	Matches:	86
Percent Similarity:	44.41%	Conservative:	41
Best Local Similarity:	30.07%	Mismatches:	123
Query Match:	14.06%	Indels:	36
DB:	14	Gaps:	14

US-10-007-527a-2 (1-379) x US-10-261-481-6 (1-2401)

QY 60 AsnGlyProLysGlySerGlyPheGlyGlyLeuArgSerGlyLysGlyTyrPheCys 79
DB 1577 GATGGCGGCGGATGCGGATCTTTTCAGCTAGCAGAGTGGGGGCTGC 1518
QY 80 ProCysGlyAlaGlyLysValGlyAlaHisArgAlaAspGluIleSerGlnValAla 99
DB 1517 GCAGTTGTAGCGCTGCATCGCGCAAAATTCGGCTTGAAAGATCATCTTGC 1458
QY 100 HisGlnLeuGlyThrGlySerVal---AlaMetValThrMetThrMetArgHisThrAla 118
DB 1457 TGGCGCGCGCAGCAACGGGTTTGGCGGCTCATTACTGTGACAGCCAGCAAGCA 1398
QY 119 GlyGlnArgLeuHisAspLeuTyrPheGlyLeuSerAlaAlaTyrLysAlaAlaThrAsn 138
DB 1397 GGTGATTCGCTTTTATGATCTTCCAAAACATGAGAGGCGAAACAGCGTTCAG 1338
QY 139 GlyArgArgTyrPheArgThrGluArgGluMetTyrGlyCysAspGlyTyrValArgAlaVal 158
DB 1337 CGCGCGAGTGGCGGATTTACATTTGTT-----GGTCAAGTCAATCACT 1290
QY 159 GluIleThrHisGly--LysAsnGlyTyrHisValHisValHisAlaLeu--LeuMet 176
DB 1289 GAAATTACTCATAGCTATGCAAAATGGGTGGCACCAGCACTTTCAGAAATGTCTGCTA 1230
QY 177 PheSerGlyAspValSerGlu-----AsnIleLeuGluSerPheSerAspAlaMetPhe 194
DB 1229 AGAGCTGGCGATGAAGCGAAGCGCTGCATCTGATGCAACGTCTAGCGCATGCG----- 1176
QY 195 AspArgTyrPheSerLysLeuValSerLeuGlyPheAlaAlaProLeuArgAsnSerGly 214
DB 1175 -----TGGCGTGCATGTCTGAAGGGGTATGGCATGTGGGCG-----AACGATGCG 1131
QY 215 GlyLeuAspValArgLysIleGlyGlyGluAlaAspGlnValIleuAlaAlaTyrLeuThr 234
DB 1130 GCGTTGACGTTGGCGGGCTGCCAATGCCGGCGAC-----TACGTAGCG 1086
QY 235 LysIleAlaSerGlyValGlyMetGluValGlySerGlyAspGlyLysSerGlyArgHis 254
DB 1085 AAG-----TGGGGCGCTGCCAAGAGGTGACGCTTCAAGCTCGAAGACGGGAAACGC 1032
QY 255 GlyAsnArgAlaProTyrGluIleAlaValAspAlaValGlyGlyAspProGlnAlaLeu 274
DB 1031 AAAGAGCGACGCGCAAGCGAGTACTT-----CAGCGGCGCATGAC----- 990
QY 275 GluLeuTyrPheArgGluPheGluPheGlySerMetGly-----ArgArgAlaIleAlaTyr 292
DB 989 GGGCTTTGGCTCGAATTTTCAATGCAACAGCGGCAAGCGGCAATGTGTGTG 930
QY 293 SerArgGlyLeuArgAlaArgAlaGlyLeuGlyAlaGluLeuThrAsp----- 308

DB 929 ACCCAAGCTTGAAGAAAGAAATCGGTTTG---GTGAGCTGTGATGACGATGAAGCTATG 873
QY 309 AlaGlnIleValGluGlnGluSerLafProValMetValAlaIleIleProAlaArg 328
DB 872 GGTGAAGTGAATGCAGCAGAGCAAGCGCCCTGAATTTGTCCGGAATGGGACATGAA 813
QY 329 SerTyrMetMetIleArg 334
DB 812 GGTGGAAGCAAGTGC 795

Search completed: August 22, 2003, 12:13:47
Job time : 267 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: August 22, 2003, 10:02:44 ; Search time 1920 Seconds

(without alignments)
4797.602 Million cell updates/sec

Title: US-10-007-527a-2

Perfect score: 2005
Sequence: 1 MTSVSAEHLGSKDRPPVLVS.....HHYHLPADVPPIISVAK 379

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Rgapop 6.0 , Rgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-Processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh
-Q=/cgm2_1/USPTO.spool/US10007527/runat_20082003.165152.15377/app_query.fasta_1.519
-DB=BSN -QFMT=fastaf -SUFFIX=p2n.rst -MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdd -LIST=45
-DOCALLIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pco -NORM=ext -HEADS=500 -MINLEN=0 -MAXLEN=200000000
-USER=US10007527 @CGN 1.1 2135 @runat_20082003.165152.15377 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

EST: *
1: em_estba: *
2: em_esthum: *
3: em_estlin: *
4: em_estmu: *
5: em_estov: *
6: em_estpl: *
7: em_estro: *
8: em_hic: *
9: gb_est1: *
10: gb_est2: *
11: gb_hic: *
12: gb_hic: *
13: gb_est3: *
14: gb_est4: *
15: gb_est5: *
16: em_estfun: *
17: em_estom: *
18: em_gss_hum: *
19: em_gss_hiv: *
20: em_gss_pln: *
21: em_gss_vrt: *
22: em_gss_fun: *
23: em_gss_man: *
24: em_gss_mus: *
25: em_gss_pro: *
26: em_gss_rtd: *
27: em_gss_phg: *
28: gb_gss1: *

29: gb_gss2: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
C 1	129.5	6.5	929	13	BUS19534	BUS19534 AGENCOURT
C 2	129	6.4	1648	12	BM913094	BM913094 AGENCOURT
C 3	128.5	6.4	1090	12	BM457396	BM457396 AGENCOURT
C 4	122.5	6.1	894	12	BI661991	BI661991 AGENCOURT
C 5	119.5	6.0	1661	29	CC321442	CC321442 TAM32-34D
C 6	119	5.9	950	13	BU271357	BU271357 603818506
C 7	119	5.9	1430	12	BM907785	BM907785 AGENCOURT
C 8	119	5.9	1687	12	BI224374	BI224374 602940916
C 9	118	5.9	851	13	BO714786	BO714786 AGENCOURT
C 10	118	5.9	893	29	BM554247	BM554247 pacsl-60
C 11	117	5.8	1306	13	BO936859	BO936859 AGENCOURT
C 12	117	5.8	1615	10	BF783863	BF783863 602109787
C 13	116.5	5.8	689	13	BU705875	BU705875 UI-M-FR0
C 14	116	5.8	867	12	BI411296	BI411296 602964678
C 15	116	5.8	1199	10	BM684181	BM684181 602635811
C 16	115.5	5.8	703	14	CG248284	CG248284 UI-M-FD0
C 17	115.5	5.8	786	13	BO572893	BO572893 UI-M-FD0
C 18	115.5	5.8	799	14	CA320693	CA320693 UI-M-FW0
C 19	115.5	5.8	857	14	CA750314	CA750314 UI-M-FD0
C 20	115.5	5.8	918	14	BY717387	BY717387 BY717387
C 21	115.5	5.8	2161	11	AK049163	AK049163 Mus muscu
C 22	115.5	5.8	2161	11	AK083015	AK083015 Mus muscu
C 23	115.5	5.8	2162	11	AK028287	AK028287 Mus muscu
C 24	115.5	5.8	2167	11	BC013842	BC013842 Mus muscu
C 25	115.5	5.8	2168	11	AK017609	AK017609 Mus muscu
C 26	115	5.7	696	14	BY729572	BY729572 BY729572
C 27	115	5.7	899	29	BZ551248	BZ551248 pacsl-60
C 28	114.5	5.7	930	12	BS922560	BS922560 602820270
C 29	114.5	5.7	1189	29	CC318464	CC318464 TAM32-32C
C 30	113.5	5.7	974	13	BO672428	BO672428 AGENCOURT
C 31	113.5	5.7	1414	13	BO276790	BO276790 AGENCOURT
C 32	113	5.6	932	10	BM444776	BM444776 CA_Ba002
C 33	113	5.6	1186	10	BF786108	BF786108 602110681
C 34	112.5	5.6	1279	10	BF160413	BF160413 601771374
C 35	112	5.6	1480	13	BU434837	BU434837 603220135
C 36	111.5	5.6	926	12	BM457074	BM457074 AGENCOURT
C 37	111.5	5.5	1221	29	CC297234	CC297234 CH261-134
C 38	110.5	5.5	984	12	BI655429	BI655429 603285076
C 39	110.5	5.5	1242	10	BM568031	BM568031 601341540
C 40	110.5	5.5	1528	12	BM904431	BM904431 AGENCOURT
C 41	110	5.5	1297	29	CC221328	CC221328 CH261-150
C 42	110	5.5	1297	29	AG111098	AG111098 Pan trogl
C 43	109.5	5.5	1157	13	BO960888	BO960888 AGENCOURT
C 44	109.5	5.5	1682	12	BM457385	BM457385 AGENCOURT
C 45	109	5.4	874	29	CC436734	CC436734 PUDCP391TD

ALIGNMENTS

RESULT 1
BUS19534/c
LOCUS BUS19534 929 bp mRNA linear EST 12-SEP-2002
DEFINITION AGENCOURT 10166703 NIH MGC 134 Mus musculus cDNA clone
IMAGE:6517370 5', mRNA sequence.
ACCESSION BUS19534
VERSION BUS19534.1 GI:22827060
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 929)

AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgabbs-remail.nih.gov
 Tissue Procurement: Dr. David Rowe
 cDNA Library Preparation: Invitrogen Corp
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
 Plate: L14M14097 row: 9 column: 03
 High quality sequence start: 6
 High quality sequence stop: 710.
 Location/Qualifiers

FEATURES

source 1.929
 /organism="Mus musculus"
 /mol_type="mRNA"
 /db_xref="taxon:10090"
 /clone="IMAGE:517370"
 /tissue_type="undifferentiated limb"
 /lab_host="DH10B (phage-resistant)"
 /clone_id="NIH_MGC_134"
 /note="Vector: PCWV-SPORT6.1.cdb; Site_1: EcoRV; Site_2: NotI; Cloned unidirectionally. Primer: Oligo dt. Average insert size 1.7 kb. Constructed by Resgen, Invitrogen Corp. Note: this is a NIH MGC library."
 BASE COUNT 183 a 312 c 251 g 183 t
 ORIGIN

Alignment Scores:

Pred. No.: 0.0953 Length: 929
 Score: 129.50 Matches: 91
 Percent Similarity: 31.90% Conservative: 28
 Best Local Similarity: 24.40% Mismatches: 95
 Query Match: 6.46% Indels: 160
 DB: 13 Gaps: 19

US-10-007-527a-2 (1-379) x BUS19534 (1-929)

QY 20 SerSerAspLysArgGlyLeuArgHisGluLeuArgProlys-----leu 34
 DB 768 GCTTCTTCACAAAAGCCGCTTCGCGCATCATCTT---CTGACACACCCCTTGCTTGACTTGG 712
 QY 35 GlnGlnIleThrThrSerGluThrPhe-AsnAlaCysGlyArgProIleSerGlyValAs 54
 DB 711 CAAAGAGTCACAGCTCTCAGGACATGTGAGACATGCTCACA----- 669
 QY 54 nglyValThrIleValAsnGlyProlysGlySerGlyPheGlyGlyLeuArgSerCysgl 74
 DB 669 ----- 669
 QY 74 yLysGlyTyrPILCysProCysCysAlaGlyLysValGlyAlaHisArgAlaAspGluI 94
 DB 668 -----TCG-----TGTCATGCTGCG----- 654
 QY 94 eSerGlnValValAlaHisGlnLeuGlyThrGlySerValAlaMetValThrMetThrme 114
 DB 653 -----GSAACATCTCAACCTGACACCTGATCTGTAGAGGT 619
 QY 114 tArgHisThrAlaGlyGlnArgLeuHisAspLeuTyrThrGlyLeuSerAlaAlaTrrply 134
 DB 618 A-----GGAACAGGCGCGGAGACCA-----GGCGAGCCAGAGCATGGA 578
 QY 134 sAlaAlaThrAsnGlyArgArgTyr ArgThrGlu-----ArgGluMetTyrGly-Cys 151
 DB 577 G-----CGCTGGTGAAGCAGAGCTTCTCCAGACATGATGCTGCTGT 536
 QY 152 AspGlyTyrValArgAlaValGluIleThrHisGlyLysAsnGly---TrpHisValHis 170
 DB 535 -----GGGCACTTGACATCTGGGGTACTAGG 509

QY 171 ValHisAlaLeuLeuMetPheSerGlyAspValSerGluAenIleLeuGluSerPheSer 190
 DB 508 CTCACAGCATTCCTTGAGCTCGGC----- 482
 QY 191 AspAlaMetPheAspArgTyrThrSerTyrLeuValSerLeuGlyPheAlaAlaProleu 210
 DB 481 -----TCTCAGGAGCCCGTCCCTGCGGCTCCCACTTCGCGC 446
 QY 211 ArgAsnSerGlyGlyLeuAspValArgGlyIleGlyGlyAlaAspGlnValleuAla 230
 DB 445 CCGCTTCAGGACGAGCTGAGCTCAGA-----GACGAGTCTCTGA 404
 QY 231 AlaTyrLeuThrIleAlaSerGlyVal---GlyMetGluValGlySerGlyAspGly 249
 DB 403 GGGATCTTGCCCGCGCTGTGACCAACATTAACAGTTGACTCTTGTTCGCTGACGAG 344
 QY 250 LysSerGlyArgHisGlyAsnArgAlaProTyrGluIleAlaValAspAlaValGlyGly 269
 DB 343 GGATGGGCAACAGGGGAGCCAGGCTGACGCTGAGAGGCTAGAGTACG----- 296
 QY 270 AspProGlnAlaLeuGluLeuTyrArgGluPheGluPheGlySerMetGlyArgArgAla 289
 DB 295 -----TCG-----AGGCTGGG 284
 QY 290 IleAlaTyrSerArgGlyLeuArgAlaArgAlaGlyLeuGlyAlaGlyLeuThrAspAla 309
 DB 283 CTCACCTGGGAACTGGGCTCAGGACGAACTGGGCGCAGG----- 242
 QY 310 GlnIleValGlnGluIleSerAlaProValMetValAlaIleIleProAlaArgSer 329
 DB 241 -----GTCCATCCAGGACGCCCTTACT-----CCGGCGCGC 209
 QY 330 TrpMetMetIleArgThrCysAlaProTyrValPheGlyGlnIleLeuGlyLeuValGlu 349
 DB 208 -----GACGGGCGCGGCTGGTCACT 188
 QY 350 AlaGlyAlaThrTyr-GluAsnLeuArgAspHisIleuHisTyrArgGluProAlaAlaAs 369
 DB 187 CCGGAGTTCGCTGGCGGCGCGGCGGCTTGGCGCTGCTCCGGTGAACGACGAGAC 128
 QY 369 pValArgProProlIleIleSerValArgLys 379
 DB 127 CACGAGGCCACAGGCGCTGGGCTGAGACAG 97

RESULT 2
 BM913094/c 1648 bp mRNA linear EST 12-MAR-2002
 LOCUS AGENCOURT 6613820 NIH_MGC_41 Homo sapiens cDNA clone IMAGE:5475102
 DEFINITION 5', mRNA sequence.
 ACCESSION BM913094
 VERSION BM913094.1 GI:19363473
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 1648)
 NIH-MGC <http://mgc.nci.nih.gov/>.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgabbs-remail.nih.gov
 Tissue Procurement: DCTD/DRP
 cDNA Library Preparation: Rubin Laboratory
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
 Plate: L14M1498 row: 9 column: 07
 High quality sequence stop: 263.
 Location/Qualifiers
 source 1.1648

```

/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5475102"
/issue_type="amelanotic melanoma, cell line"
/lab_host="DH10B (phage-resistant)"
/clone_1ib="NIH_MGC_41"
/note="Organ: Skin; Vector: pOTB7; Site 1: XhoI; Site 2:
EcoRI; cDNA made by oligo-dT priming. Directionally cloned
into EcoRI/XhoI sites using the following 5' adaptor:
GGCAGCAG(G). Library constructed by Ling Hong in the
laboratory of Gerald M. Rubin (University of California,
Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and
Superscript II RT (Life Technologies). Note: this is a
NIH MGC Library."
BASE COUNT      333 a      668 c      437 g      205 t      5 others
ORIGIN

```

Alignment Scores:

```

Pred. No.:      0.227      Length:      1648
Score:          129.00     Matches:      93
Percent Similarity: 30.73%  Conservative: 33
Best Local Similarity: 22.68% Mismatches:     153
Query Match:      6.43%    Indels:       131
DB:              12       Gaps:        17

```

US-10-007-527a-2 (1-379) x BM913094 (1-1648)

```

QY 10 SerGlyAspArgProValLeuValSerSerAspLysArgGlyLeuArgHisGlu 29
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1101 GCTGGCCCAAGCGCGCCGACATCAATATATACC-----CCACTGGGGCTCCGACACCGA 1048

QY 30 LeuArgProLys----- 33
   |||||
DB 1047 AACGCTCCAGCGGTCGGGCGACAGAGAGCGTGTGACTGTGCTTGAACAGATTGG 988

QY 34 ---LeuGlnGlnIleThrThrSerGluThrPheAsnAlaCysGlyArgPro----- 49
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 987 CCGTTGAGGACACTCCGACAGNAG-----GGGCGCGGTGAGACTGTA 946

QY 50 -----IleSerGlyValAsn----- 54
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 945 AGTTGCTAGAGGGGCGTTATATCTGCTGCGCTGCGGATGATGATGACGACAGAGAGAG 886

QY 55 -----GlyValThrIleValAsnGlyProLysGly 64
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 885 GTGTGATTGTGCTCGGTCGGCGGCTTAGTAGGCACAGGTGTGTGGGACCATGTGGT 826

QY 65 SerGlyPheGlyGlyLeu-ArgSerCys-----GlyLysGlyT 77
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 825 GCGGCTGTGCTGCGGCTCAGCTGCTGCGGCTGGGAGTGCCTGCTGCGGCTGTGGAT 766

QY 77 rPLeGcysProCys-----CysAlaGlyLysValGlyAlaHisArgA 91
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 765 GGGGTTGTGGTTGGCGGCAATTGGAGCGCGCTGTGTGCCCATTTGGTGGCGCGG 706

QY 91 laAspAluIleSerGlnValAlaHisGlnLeuGly--ThrGlySerValAlaMetV 110
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 705 GACCTGTTTGGCTGCTGCGGCACTAGCATGCGGCGCGGAGGTGTGTCATGGGCA 646

QY 110 alThrMetThrMetAlaGlnHisThrAlaGlyGlnArgLeuHisAspLeuTrpThrGlyLeus 130
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 645 TTTCAGCCGAGCGCGGCGCATCCCGCC-----GAGGGGCGCG 610

QY 130 eraIaIaIaTrpLysAlaAlaThrAsnGlyValArgArgTrpArgThrGlnArgGlyMetTyrG 150
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 609 GGGGGAATCTGG---GTTGCGCGGCGACAGCCCATGTGGCGCTTGGCGGCGACAGTGG 553

QY 150 lYcysAspGlyTyrValArgAlaValGluIleThrHisGlyLysAsnGly--TrpHisVal 169
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 552 GGGGCGGTGGGC-----GGTAGTGGCGGTTTGGCACTCC 520

QY 170 HisValHisAlaLeuLeuMetPheSerGlyAspValSerGluAsnIleLeuGlnUserPhe 189

```

```

DB 519 GAGACGGGGGGGATGCTGTGACTGCTGG----- 490
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 190 SerAspAlaMetPheAspArgTrpThrSerLysLeu----- 201
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 489 -----TGAACGGCGGCTCAGCGAGGATTCGTGGCGGGCTCCGGCT 451

QY 202 ---ValSerLeuGly--PheAlaAlaProLeuArgAsnSerGlyGlyLeuAspValArgLys 220
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 450 GGTGTGGCGGTTGGCGCTTGAACCGCGGCGCGCTATAGTGGACAGGTCGCGGCTTTC 391

QY 220 sIlleGlyGlyGlnAlaAspGlnValLeuAlaAlaTyrLeuThrLysIleAlaSerGlyVa 240
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 390 TGTGGCGGTGCTGTTGAGTGAAGTTGCGGGGGGTGTGCACACAGATTTTGTGC----- 336

QY 240 lGlyMetGluValGlySerGlyAspGlyLysSerGlyArgHisGly-----As 256
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 335 -GACATGTGTGCGGGCGCGGCGCGGCGCTGTGTGGGAGCCGGGTCCTCCGACCGAGAC 277

QY 256 nArgAlaProTrpGluIleAlaValAspAlaValGlyLysProGlnAlaLeuGluLe 276
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 276 GAGAGGGGCGCTGG-----ACGGCGAGGGGTGGCGCGCGGACCACTTGCGG 229

QY 276 uTrpArgGluPheGlu-PheGlySerMetGlyArgArgAlaIleAlaTrp----- 292
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 228 GTGGCGCGCGCGCGCGACAGCGGCGCGCGCGCTGTGCGCAACTGGTCTCTGGA 169

QY 293 -----SerArgGlyLeuArgAlaArgAlaGlyLeuGlyValGluLeuThrAspAlaGlnI 311
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 168 GAGCGCGGACAGAGCGCGCGGGTGGGGCGGGGTGTGTGAGCGGAGACCGGACAGGCG 109

QY 311 lValGlnGlnGlnGlnGlnSerAlaProValMetValAlaIleIleProAlaArgSerTrpM 331
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 108 TCGGCGCTTCCGAGAGCGCGCGGAGAGGCGGCTGACAGCGGTGCGCGCGCTCGGCGG 49

QY 331 eMetIleArgThrCysAla 337
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 48 CCGTAGCTCATCTCCGGGCT 29

```

```

RESULT 3
BM457396      1090 bp      mRNA      linear      EST 05-FEB-2002
LOCUS
DEFINITION
AGENCOUNT 6407558 NIH_MGC_92 Homo sapiens cDNA clone IMAGE:5584861
5', mRNA sequence.
ACCESSION
BM457396
VERSION
BM457396.1 GI:18506436
KEYWORDS
SOURCE
ORGANISM
Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

```

```

REFERENCE
1 (bases 1 to 1090)
NIH-MGC http://imgc.ncl.nih.gov/.
AUTHORS
National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE
Unpublished
JOURNAL
Contact: Robert Strausberg, Ph.D.
Email: cgaabs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LHAM12350 row: d column: 14
High quality sequence drop: 538.
Location/Qualifiers

```

FEATURES

source

```

1..1090
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5584861"
/tissue_type="embryonal carcinoma, cell line"

```

/lab_host="DH10B (phage-resistant)"
 /clone_lib="NIH_MGC_92"
 /note="Organ: testis; Vector: pCMV-SPORT6; Site 1: NotI;
 Site 2: SalI; Cloned unidirectionally; oligo-dT primed.
 Average insert size 2.5 kb. Library enriched for
 full-length clones and constructed by life technologies.
 Note: this is a NIH_MGC library."

BASE COUNT 286 a 250 c 398 g 156 t

ALIGNMENT SCORES:

Pred. No.: 0.146 Length: 1090
 Score: 128.50 Matches: 80
 Percent Similarity: 31.38% Conservative: 27
 Best Local Similarity: 23.46% Mismatches: 94
 Query Match: 6.41% Indels: 140
 DB: 12 Gaps: 15

US-10-007-527A-2 (1-379) X BM457396 (1-1090)

```

QY 7 GJUHLEU-----SerGlyAspArgProPheValLeu-----18
DB 314 GAGCAGCTGACGACTGAGCCAGAGAGGTGCCAGAGTGTAAAGCTGTGCAGAAATT 373
QY 19 -----ValSerSerAspLys 23
DB 374 GTGAGAGATGATGAGATGGATGGATCTACCGCTCTCAGGGGTCTCTCCACATTC 433
QY 24 ArgGlyTLeArgHisGluLeu-----ArgProLysLeuGlnGln-1LeuThr 39
DB 434 CAGAACCTTCGGAGAGAAATTGAGTCAGACCGAGCCAGACCTGGTGGGTGTTTAC 493
QY 39 rSerGluThrPheAsnAla-----CysGlyAspProLysSerGlyValAsnGlyLys 56
DB 494 CTCCAGACACTTACCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 539
QY 56 lThrIleValAsnGlyProLysGlySerGlyPheGlyGlyLeuArgSerCysGlyLysG 76
DB 540 -----AAGAA 544
QY 76 YTPRIleCysProCysCysAlaGlyLysValGlyAlaHisArgAlaAspGluLysSerG 96
DB 545 CTCCCGGATCCCTGCG-----TCACCTTACCGGGCTCTAT-----GACAAAGATTGCTGA 592
QY 96 nValValAlaHisGlyLeuGlyThrGlySerValAlaMetValThrMetThrMetArgHis 116
DB 593 AGCTGATGAGATGGAATGGAACCTGAGCCTTGCTGACACATCTTGAGGTGCTTGGGA 652
QY 116 sThrAlaGlyGlnArgLeuHisAspLeuTrpThrGlyLeuSerAlaAlaTrp-----133
DB 653 ACTCCCTGTCGCAAACTACAGAAACCTGAGATCCTGATGAGGCACTTGTGTAACA 712
QY 134 -----LysAlaAlaThrAsnGlyArgArgTrpArg-----143
DB 713 TGGCCCTCATTGGGGGACACAGACCAAGGGTGCAATGTCACGACCTAGGGAAGAGG 772
QY 144 -----ThrGlu-----ArgGluMetTrpGlyCysAspGlyTrpVal 155
DB 773 GGGGGGTACAGAACCCAGCTAAAGGGGATAAAGAAACAGAAAGATGCCGGGGGTTCAA 832
QY 155 lArgAlaValGluIleThrHisGlyLysAsnGlyTrpHisValHisValHisAlaLeuLe 175
DB 833 AAAGGAAAGCCGGGGCACACAGGGAAGAAAGAGGGGGC-----869
QY 175 uMetPheSerGlyAspValSerGluAsnIleLeuGluSerPheSerAspAlaMetPheAs 195
DB 869 -----869
QY 195 pArgTrpThrSerLysLeuValSerLeuGlyPheAlaAlaProLeuArgAsnSerGlyG 215
DB 870 -----GGTCTGGCC-----AAGAAAAAGGGGGG 892
QY 215 yLeuAspValArgLysIleGlyGlyGluAlaAspGlnValLeuAlaAlaTrpLeuThr 235

```

```

DB 893 AAACAAAGCGCGGAAAAAGCGGGGAAAA-----920
QY 235 sIleLeaSerGlyValGlyMetGluValGlySerGlyYAspGlyLysSerGlyArgHisG 255
DB 921 -----GGGGCGGAAAAAGGGCGGGAAGGGGGCGGAAAAAGAAAGGGCGGG 970
QY 255 yAsnArgAlaProTrpGluIleAlaValAspAlaValGlyGlyAspProGlnAlaLeuG 275
DB 971 G-----GTCCCGGAAAAAAACAAAGCGGGCGGAAAAAGGG---GA 1009
QY 275 lLeuTrpArgGluPheGluPheGlySerMetGlyYArgArgAlaIleAlaTrpSerArg 295
DB 1010 ACAAAGAAAGGGGGCGGAAACAGAGCGGAAAGGGCGGCAACGGGGGTAAAGGGG 1069
QY 295 Y 295
DB 1070 G 1070

```

RESULT 4

BI661991/c 894 bp mRNA linear EST 12-SEP-2001
 LOCUS 60330498661 NC1_CGAP_Mam4 Mus musculus cDNA clone IMAGE:5350797 5',
 DEFINITION mRNA Sequence.

ACCESSION BI661991 GI:15576227

VERSION BI661991.1 GI:15576227

KEYWORDS EST.

SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus

REFERENCE Eukaryotes; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

AUTHORS NIH-MGC http://mhc.nci.nih.gov/.

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished

COMMENT Contract: Robert Strausberg, Ph.D.

Email: cgabbs-remail.nih.gov

Tissue Procurement: Lothar Hennighausen Ph.D., Priscilla Furch

Ph.D.

CDNA Library Preparation: Life Technologies, Inc.

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: LLNL1891 row: k column: 22

High quality sequence stop: 849.

Location/Qualifiers

1. 894

/organism="Mus musculus"

/mol_type="mRNA"

/strain="NMRI"

/db_xref="taxon:10090"

/clone_type="IMAGE:5350797"

/tissue_type="tumor, gross tissue"

/dev_stage="5 months"

/clone_lib="NIH_MGC"

/lab_host="DH10B"

/note="Organ: mammary; Vector: pCMV-SPORT6; Site 1: SalI;
 Site 2: NotI; Cloned unidirectionally. Primer: Oligo dT.
 Library constructed by Life Technologies. Investigators
 providing samples: Lothar Hennighausen/Priscilla Furch,
 NIH Reference for transgenic model: Li et al., Cell Growth
 and Differentiation 7, 3-11 (1996)."

BASE COUNT 176 a 305 c 243 g 170 t

ALIGNMENT SCORES:

Pred. No.: 0.398 Length: 894
 Score: 122.50 Matches: 79
 Percent Similarity: 30.77% Conservative: 25
 Best Local Similarity: 23.37% Mismatches: 114
 Query Match: 6.11% Indels: 120

DB: 12 Gaps: 16

US-10-007-527a-2 (1-379) x B161991 (1-894)

QY 68 GlyGlyLeuArgSerCysGlyGlyGly-----Trp-----IleCysPro 80
 DB 811 GGGACATTGACATCTCTCGGAGACGATACAGCTGGGATACAGGTTTTCATTTCCTCCCA 752
 QY 81 CysCysAla-----GlyLysValGlyAlaHisArgAlaAspGluIleSer 95
 DB 751 CATTGGCTCTTCAAGACGCTTGCAGATCATCTCTCGAGACACTTGACCTTGAC----- 701
 QY 96 GlnValAlaAlaHisGlnLeuGlyThrGlySerValAlaMetValThrMetThrMetArg 115
 DB 700 -----TTTGGCAAGGTCACAGCTCAGACGATTTGGGAGCATGCT 659
 QY 116 HisThr-----AlaGlyGlnArgLeu-----HisAspLeuTrpThrGlyLeuSer 130
 DB 658 CACAAAGGTGTCATGCTCGGACATCTTCACATCTGACTTGTAGGGTTCAGAC 599
 QY 131 AlaAlaTrpLysAla-----AlaThrAsnGlyArgArgTrpArg 143
 DB 598 AAGGCTGGGAGACAGGCTGAGCAGACAGATGAAAGCGCTGTAGGACAGGACTTTC 539
 QY 144 ThrGluArgGluMetArgGlyCysAspGlyTrpValArgAlaValGluIleThrHisGly 163
 DB 538 GACACATGTTGCTGTGGGACCTTGGCATC----- 506
 QY 164 LysAsnGlyTrpHisValHisValHisAlaLeuLeuMetPheSerGlyAspValSerGlu 183
 DB 505 -----TGGGACTAGAGCTCCACAGATCTCTTGACCTGGGC----- 467
 QY 184 AsnIleLeuGluSerPheSerAspAlaMetPheAspArgTrpThrSerLysLeuValSer 203
 DB 466 -----TCTCAGGGCCCGTGC 452
 QY 204 LeuGlyPheAlaAlaProLeuArgAsnSerGlyGlyLeuAspValArgLysIleGlyGly 223
 DB 451 CTGGGCTCCCACTTCCTGCGCCCGCTCAGCAGGAGCTGAGAGCTCAGA----- 404
 QY 224 GluAlaAspGlnIleLeuAlaAlaIleValTrpLeuThrLysIleAlaSerGlyVal---GlyMet 242
 DB 403 -----GACGAGTCTCTGAGGGGATCTTGGCGGTGTGCAACATTAACAGTTGT 350
 QY 243 GluValGlySerGlyAspGlyLysSerGlyArgHisGlyAsnArgAlaProTrpGluIle 262
 DB 349 ACTCTTGTTCGCTTACCGGGGATGGGCAACAGGGGAGCCAGGCTCAGCTGAGGGCT 290
 QY 263 AlaValAspAlaValGlyGlyAspProGlnAlaLeuGluLeuTrpArgGluPheGluPhe 282
 DB 289 AGGCTCAGC-----TGG----- 278
 QY 283 GlySerMetClyArgArgAlaIleAlaIleAlaTrpSerArgGlyLeuArgAlaArgAlaGlyLeu 302
 DB 277 -----AGGCTGGGCTCAGCTGGGGAATCTGGGCTCAGGCAAGAACTGGAGGCA 230
 QY 303 GlyAlaGluLeuThrAspAlaGlnIleValGluGlnGluIleSerAlaProValMetVal 322
 DB 229 GGA-----GTCCATCCAGGAGCCGTAACCT----- 203
 QY 323 AlaIleIleProAlaArgSerTrpMetMetIleArgThrCysAlaProTrpValPheGly 342
 DB 202 -----CCGGCGCGC----- 194
 QY 343 GluIleLeuGlyLeuValGluAlaGlyAlaThrTrp---GluAsnLeuArgAspHisIleuH 362
 DB 193 GACGAGCGCGGGCTGGTCACTCCCGGAGTTCGGCTGGGGCGCGGCTTGGCGGCTGCT 134
 QY 362 sTrArgLeuProAlaAlaAspValArgProPoleIleSerValArgLys 379
 DB 133 GCTCCGGGTGAAGCAGAGCCACAGGCCACAGGCTCGGGGTGAAGAC 82

RESULT 5

CC321442/c

LOCUS CC321442 1661 bp DNA linear GSS 14-MAY-2003

DEFINITION TAM32-34D5_ECI.1 TAM32 Gallus gallus genomic clone TAM32-34D5, genomic survey sequence.

ACCESSION CC321442

VERSION CC321442.1 GI:30715500

KEYWORDS GSS.

SOURCE

ORGANISM Gallus gallus (chicken)

Galus gallus

Bukarjota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.

1 (bases 1 to 1661)

REFERENCE

AUTHORS Kremetzki, C., Higginbotham, J., Wylie, K., Carter, J., McPherson, J., Warren, W., Graves, T., Mardis, E. and Wilson, R.

TITLE Gallus gallus BAC End Reads

JOURNAL Unpublished

COMMENT

Contact: Richard K. Wilson

Genome Sequencing Center

Washington University School of Medicine

Email: submissions@watson.wustl.edu

Insert Length: 15200 Std Error: 0.00

Seq primer: ECI TACGACTCATATAGGGCG

Class: BAC ends

High quality sequence start: 74

High quality sequence stop: 124.

Location/Qualifiers

1..1661

/organism="gallus gallus"

/mol_type="genomic DNA"

/strain="Red Jungle Fowl"

/db_xref="taxon:9031"

/clone="TAM32-34D5"

/sex="female"

/cell_line="UCD001, Inbred 256"

/note="Vector: pBCEBAC1; Site 1: EcoRI, Site 2: EcoRI; TAM32 Female Chicken library - for library and clone ordering information: http://www.hbz.tamu.edu"

BASE COUNT 311 a 949 c 157 g 238 t 6 others

ORIGIN

Alignment Scores:

Pred. No.: 1.71 Length: 1661

Score: 119.50 Matches: 77

Percent Similarity: 33.23 Conservative: 31

Best local Similarity: 23.69 Mismatches: 118

Query Match: 5.968 Indels: 99

DB: 29 Gaps: 14

US-10-007-527a-2 (1-379) x CC321442 (1-1661)

QY 54 AsnGlyValThrIleValAsnGlyProLysGlySerGlyPheGlyGlyLeuArg----- 71
 DB 1492 GATGGCTGACGGAGTCCGTGTGAGAGGGGAGTGATCAGAGGCCAGAGGGGTTC 1433
 QY 72 -----Ser 72
 DB 1432 ATGTGGAGGAGGAGGAGGAGGAGGAGGCTTGTGAGGAGGAGGAGTATATGAGGAGT 1373
 QY 73 CysGlyLysGlyTrpIleCysProCysCysValGlyLysValGlyAlaHisArgAlaAsp 92
 DB 1372 GTGGGAGAGGTAAATAGGTGT-----GTGAGTTCGGGAGT 1337
 QY 93 GluIleSerClyValValAlaHisGlnLeuGlyThrGlySerValAlaMetValThrMet 112
 DB 1336 ACGAGAGGTGTTGTGTGTGTCATGATGTAGGGCGAGGGGCGGGCTGAGGTGGCTCG 1277
 QY 113 ThrMetArgHisThrAlaGlyGlnArgLeu-HisAspLeuTrpThrGly----- 128
 DB 1276 GGGGCTGAGCTGTGAGGGGGGCAACGGGTAGAGG-----TGGGAGCTGGGGGTGGAG 1223
 QY 129 ----LeuSerAlaAlaTrpLys----AlaAlaThrAsnGlyArgArgTrpArgThrGluArg 146

Db 562 -----TGTGTGGGGCTGG 550
QY 207 laAlaProLeuAaGaaSerGlyGlyLeuAspValArgLysIleGlyGlyAlaAspG 227
Db 549 TTTCAGAGCTGCGCTGGGTTTGGTGGCTCCGGGTTTGGCTGGAGGTTGGCTTGGG 490
QY 227 InVal-----LeuAlaAlaTyrLeuThrLysIleAlaSerGlyValG 241
Db 489 CCGTGAAGTTCAGTGGCTGGGGCCATCTCAGGTAGCTTGTGTGAGCTTGGGATCGA 430
QY 241 lMetGluValGlySerGlyAspGlyLysSerGlyArgHisGlyAsnArgAlaProTyrG 261
Db 429 GCGTGAGCAGGGGGTGTGA-----TGGC 406
QY 261 lulleAlaValAspAlaValAlGlyLysProGlnAlaLeuGluLeuTyrPargLuhpeg 281
Db 405 GAGTGCAGGTAGAGAGCAAGTCTGGCAA----- 377
QY 281 luhpegLysMetGlyArgArgAlaIleAlaTyr-----SerArgGlyLeuArgAla 299
Db 376 -----GGCCCCGAGGCAAGGCCCGTAGGGCTTGGTTATCGTTGCATGGGATCGAGCA 322
QY 299 rGlaGlyLeuGlyAlaGluLeuThrAspAlaGlnIleValGluGlnGluGluSerAlaP 319
Db 321 GAGCA-----TCAGCCG 310
QY 319 roValMetValAlaIleIleProAlaArg-----SerTyrMetMetIleA 334
Db 309 CGGGGCTCAGAGTTGCTTCCACACGAGCTTCAGATGAGCTTCATGGCTTCAGAGCA 250
QY 334 rGhrCysAlaProTyrValPheGlyGluLeuGluLysLeuValGlu----- 349
Db 249 CAAGAAGCTGAGTTGACGTTGAGGCAAGATCTGTGGGCGCCAGCAAAACACAGAGCA 190
QY 350 --AlaGlyAlaThrTyrGluAsnLeuArgAspHisLeuHisTyrArg-----LeuP 366
Db 189 GGGCTGGACCATCAAGATAGCATTCCTTCAGCCAGCATTTAAAGCCCTGCTGCTCC 130
QY 366 roAlaAlaAspValArgProPro 373
Db 129 CTGACAGCATG-GAGCGTCCGCT 108

RESULT 7
BM907785 1430 bp mRNA linear EST 12-MAR-2002
LOCUS BM907785.1
DEFINITION AGENCOURT 6618957 NIH_MGC_68 Homo sapiens cDNA clone IMAGE:5736201
5', mRNA sequence.
ACCESSION BM907785
VERSION BM907785.1 GI:19358164
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 1430)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabs-r@mail.nih.gov
Tissue Procurement: DCTD/DTF/Gasdar
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.llnl.gov
Plate: L1AM12744 row: f column: 10
High quality sequence stop: 319.
Location/Qualifiers
1..1430
/organism="Homo sapiens"
/mol_type="mRNA"

/db xref="taxon:9606"
/clone="IMAGE:5736201"
/tissue_type="large cell carcinoma"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH MGC 68"
/note="Organ: lung; Vector: pCMV-SPORE; Site 1: NotI;
Site 2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.8 kb. Library constructed by Life
Technologies."

BASE COUNT 341 a 598 c 228 g 258 t 5 others
ORIGIN

Alignment Scores:
Pred. No.: 1.56 Length: 1430
Score: 119.00 Matches: 90
Percent Similarity: 34.82% Conservative: 43
Best Local Similarity: 23.56% Mismatches: 110
Query Match: 5.94% Indels: 139
DB: 12 Gaps: 21

US-10-007-527A-2 (1-379) x BM907785 (1-1430)

QY 46 CysGlyArgProIleSerGlyValAsnGlyValThrIleValAsnGlyProLysGlySer 65
Db 1343 TGTGGGCGCCCA-----GTTATGGGGGGGTATTGG-----GGGGGTGGGGGAAA 1299
QY 66 GlyPheGlyGlyLeuArgSerCysGlyLys----- 75
Db 1298 GGTGTGGGTGTATGTTTATATGTGTGTCCTCCGCTTATGTGAGGGGGGGGGGG 1239
QY 76 -----GlyTyrIleCys 79
Db 1238 GTGGATATGAGATGGGGGGGTACGTCCTCATATGACAGGGCGGTGTATTATTGG 1179
QY 80 ProCys-----CysAlaGlyValValGlyAlaHisArgAlaAsp 92
Db 1178 CAGTGTGGCGAGATATGGGTATTATCGGGGGCGATTCCGA-----AGGTGAAGN 1125
QY 93 GluIleSerGlnValAlaHisGlnLeuGly-ThrGlySer-----ValAlaMetVal 110
Db 1124 GAGCTCTTATGCTGTATATGCGGATATCGGCATATGAGGCGTTGATGATGAGCTCGT 1065
QY 110 lThrMetThrMetArgHisThrAlaGlyIleArgLeuHisAspLeuTyrThrGlyLeu-- 129
Db 1064 GACTATCGCT--AAACAT-----TTGTGGGTGGTGTGCTG 1032
QY 130 -----SerAlaAlaTyr-----Lys-AlaAlaThrAsnG 139
Db 1031 CCGGACCGGGGAGTCAGTGGGCGCTTATGCGGTTTGTCCGCGTGGGATATACCTGTG 972
QY 139 lArgArgTyrPargThrGluArgGluMetTyrGlyCysAspGlyTyrValArgAlaValG 159
Db 971 GTAGCGGGCGCCCGGAGAAAGAGGCAAGTTAAAGTGACGGG-----GCTTTAT 921
QY 159 lulleThrHisGlyLysAsnGlyTyrHisValHisValHisAlaLeuLeuMetPheSerG 179
Db 920 GTCTAAGCTAAGGCTTATAGAGGGGTGGCGAAGGGTGTGACGCTTTTGTGGTGG 861
QY 179 lAspVal-----SerGluAsnIleLeuGluSerPheS 190
Db 860 GCGGCGTGAAGATATGTCGCGCGGTATAGAGAAATCCAGATTTGGATGTCCGCAATT 801
QY 190 exAspAlaMetPheAspArgTyrThrSerLysLeuValSerLeuGlyPheAlaAlaProL 210
Db 800 CGAACAGTAT-----CTCAGCCTTATATATGCG-----G 771
QY 210 euArgAsnSerGlyGlyLeu-----AspValA 219
Db 770 TAGCGGAATCAGCGCGCTGTATATGACTCATTAAGGTTGCCGANAAGTCATGAGTGTG 711
QY 219 rGlySerIleGlyGly-----GluAlaAspGlnValLeuAla 231
Db 710 GCGTATAGGTGTATTTTGGCCAGTATGTGAGGTGGAGGGGTATTAAGAGATGGGGA 651

```

QY      231 laTyrLeuThrLysIleAlaSerGlyValGlyMet----- 242
      650 ::::|||||:
      Db      650 GTATAGGCTGCTTGTGAATCAAGGAGAGTTTGCGGTAGCGCATGACGTAAATGCG 591
QY      243 -----GluValGlySerGlyAspGlyLysSerGlyAAspGlyAsnArgAlaProTyrG 261
      590 CCGAGGGGCCAGAGGTTTGGCGAGGAGATAGGGGTGATGTGGTGGCAGCGCTCCCTTTT 531
      Db      261 luilealaValAspAlaValAlaGlyLysProGlnAlaLeuGluLeuThrArgGluPheG 281
      530 GGCATAGGCTGCGGAGCGCTTGT-----GTTGAGTGGCTCAGAGGATTTTG 466
QY      281 lupheGlySerMetGlyArgArgAlaIleAlaIleArgSerArgGly-----LeuA 297
      485 GGGAGAGGTAGTGGTGAATGCGCTGTATTGGGCGCATAGAGGGGATGATGCTTTGGC 426
      Db      297 rg-AlaArgAlaGlyLeuGlyAlaGluLeuThrAspAlaGlnIleValGluGlnGlu 316
      425 GTGGGCGAGGGGGGCGAGGGGCG-----GCGGGGGGTGGTGGGGCGAGGT--- 381
QY      317 SerAlaProValMetValAlaIleIleProAlaArgSerTyrMetMetIleArgThrCys 336
      380 -----CCGGGGCGTGGGTGGTGGTGTACTAGAGTGTGT 348

```

```

RESULT 8
LOCUS   BI224374 1687 bp mRNA linear EST 11-JUL-2001
DEFINITION 602940916F1 NIH_MGC_12 Homo sapiens cDNA clone IMAGE:5104346 5',
mRNA sequence.
ACCESSION BI224374
VERSION   BI224374.1 GI:14677818
KEYWORDS  EST.
SOURCE    Homo sapiens (human)
ORGANISM  Homo sapiens

```

```

REFERENCE 1 (bases 1 to 1687)
AUTHORS   NIH-MGC http://mgs.nci.nih.gov/.
TITLE      National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL    Unpublished
COMMENT    Contact: Robert Strausberg, Ph.D.
           Email: cgabs-remail.nih.gov
           Tissue Procurement: ATCC
           cDNA Library Preparation: Life Technologies, Inc.
           cDNA Library Arrayed by: Incyte Genomics, Inc.
           DNA Sequencing by: Incyte Genomics, Inc.
           Clone distribution: MGC clone distribution information can be
           found through the I.M.A.G.E. Consortium/BLMT at:
           http://image.llnl.gov
           Plate: LHAM11251 row: 0 column: 03
           High quality sequence stop: 195.

```

```

FEATURES
Source    Location/Qualifiers

```

```

1..1687
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5104346"
/tissue_type="cervical carcinoma cell line"
/lab_host="DH10B"
/clone_id="NIH_MGC_12"
/note="Organ: Cervix; Vector: PCMV-SpOR6; Site 1: NotI;
Site 2: SalI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 1.4 kb. Library prepared by Life
Technologies."

```

```

BASE COUNT      286 a      842 c      254 g      305 t
ORIGIN

```

```

Alignment Scores:
Pred. No.:      1.94      Length:      1687
Score:          119.00     Matches:      81
Percent Similarity: 33.72%  Conservative: 34
Best Local Similarity: 23.75% Mismatches:     102

```

```

Query Match:      5.94%      Indels:      124
DB:               12       Gaps:      18
US-10-007-527A-2 (1-379) x BI224374 (1-1687)

```

```

QY      61 GlyProLysGlyIleSerGlyPheGlyLysLeuArgSerCysGlyLysGly-----Tyr 77
      1036 GGGAGGGTGGGAGGTGAGCGAGCGAGTGG-----GGCTGGCGGAGAGGCTGTCAGGTGG 983
      Db      78 IleCys-----ProCysAlaGlyLys-----ValGlyAlaHisArg 90
      982 CTGTGATGAGGGGGGAGTGTGTGGCGGAGCGCGCGCGCGGAGATGGTGTGCGCCGC 923
QY      91 AlaAspGluIleSerGlnValAlaAlaHisGlnLeuGlyThrGlySerValAlaMetVal 110
      922 GCGCGAGAGGTG---AAGGAGGGGCGAGCGCGGAGCGGAGGATGAGAGCG----- 878
      Db      111 ThrMetThrMetCargHisThrAlaGlyGlnArg---LeuHisAspLeuThrGlyLeu 129
      877 -----GCTGGAGCACGCTGAGAGAGAGACCTGTGTGGGGGGGCTG 839
QY      130 SerAlaAla-TripLysAlaAlaThrAsnGlyArgArgTyrArgThrGluArgGluMetC 149
      838 ACATGAGCACTGCGCGAGGGCGGAGAGCGGTGCGCTGAGAGAGC----- 793
      Db      149 rGlyCysAspGlyTyrValArgAlaValAlaIleThrHisGlyLysAsnGlyTyrHisVa 169
      792 -----AGGGGGTGGCGA--- 781
QY      169 HisValHisAlaLeuLeuMetPheSerGlyAspValSerGluAsnIleLeuGluSerPh 189
      780 -----CATGCT----- 775
      189 eSerAlaIleMetPheAspArgTyrThrSerLysLeuValSerLeuGlyPheAlaIleApr 209
      774 -----GGCAACGCGCGCC 761
QY      209 OleuArgAsnSerGlyGly----- 215
      760 GCCAGAGCAACCGCGGAGAGCGCGCGGCGGAGAGAGAGAGAGAGAGAGTGGG 701
      Db      216 -----LeuAspValArgGlySerIleGlyGlyGlu-----AlaAspGlnValLeuAl 230
      700 GAGGCCAGTGTGGAGCCGCGGCGGCGGAGAGAGTGGGTGACAGATGTGCGGC 641
QY      230 AlaTyrLeuThrLysIleAlaSerGlyValGlyMetGluValGlySer----- 246
      640 G-----GAGAGCGCGCGGGGTGTGACAGAGAGACCCGTGGAGAA 599
      Db      247 ---GlyAspGlyLysSerGlyArgHisGlyAsnArgAlaProTyrGluIleAlaValAs 265
      598 GATGGGTGAGGCGGAATGATGACACAGCGAGCTGGCCGAGCGAGGGGTG 539
QY      265 palAval-----GlyGlyAspProGlnAlaLeuGluLeuThrArgGluPh 280
      538 GTTGACAGAGGCTGTGCGAGAGGGGCAACGGGCGGAA-----TGGAGCGCGG 488
      Db      280 eGluPheGlySerMetGlyArgArgAlaIleAlaIleTyr-----SerArgGlyLeuArgAl 298
      487 GCGCCGGGGGCGTGAAGGAGAGGGGGCGAGATGCTGTGACGACGATGTGTGCGCGC 428
QY      298 ArgArgAlaGlyLeuGlyAlaGluLeuThrAsp-AlaGlnIleValGluGlnGluSerA 318
      427 AGCTAGGGGCGCATGTGCGCGAGATGCGCGAGGGGCGGAGCACGCGAGGGCGAGAG 368
      Db      318 laProValMetValAlaIleIle-----ProAlaArgSerTyrMetMetIleArgT 335
      367 CGCGCTGGGAGCGAGCGAGGCAAGGGCTGAGCTGCGAGCGGCGACTGTGTGCGGAG 308
QY      335 hrcysAlaProTyrValPheGlyGluIleLeuGlyLeuValGluAlaGlyAlaThrTyr 354
      307 TGTGCAAGCGAGCGGTGCCAGCAGGTGCTGTGGTTCCATGGGGAATTAAGCTGTGTG 249

```

RESULT 9
 B0714786/c
 LOCUS B0714786 851 bp mRNA linear EST 16-UTL-2002
 DEFINITION AGENCOURT 8482181 NIH_MGC_129 Mus musculus cDNA clone IMAGE:6307476
 5', mRNA sequence.
 ACCESSION B0714786
 VERSION B0714786.1 GI:21853685
 KEYWORDS EST.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE NIH-MGC <http://mgc.nci.nih.gov/>.
 1 (bases 1 to 851)
 AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgabbs-remail.nih.gov
 Tissue Procurement: Susan L. Sullivan, PhD.
 cDNA Library Preparation: Resgen, Invitrogen Corp.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
 Plate: LHM3725 row: a column: 13
 High quality sequence stop: 666.
 Location/Qualifiers
 1..851
 /organism="Mus musculus"
 /mol_type="mRNA"
 /db_xref="taxon:10090"
 /clone="IMAGE:6307476"
 /lab_host="DH10B (phage-resistant)"
 /clone_lib="NIH MGC 129"
 /note="Organ: olfactory epithelium; Vector:
 pCMV-SPORT6.1.ccd; Site_1: EcoRV, Site_2: NotI, Cloned
 unidirectionally. Primer: Oligo dt. Average insert size
 2.2 kb. Constructed by Resgen, Invitrogen Corp. Note: this
 is a NIH MGC Library."
 BASE COUNT 164 a 289 c 234 g 164 t
 ORIGIN
 Alignment Scores:
 Pred. No.: 0.964 Length: 851
 Score: 118.00 Matches: 78
 Percent Similarity: 32.43% Conservative: 30
 Best Local Similarity: 23.42% Mismatches: 112
 Query Match: 5.89% Indels: 113
 DB: 13 Gaps: 15
 US-10-007-527a-2 (1-379) x B0714786 (1-851)
 QY 68 GYGYLYLWARGSERGSLYSGLYTPTLIECYSPROCYCYALAGLYLYVALGY 87
 DB 845 GCGGAAAGCGGATACGACCGTGGGGATAC-----ACCGGTTTGATTGG 801
 QY 88 ALAHISARGALASPGLIULESERGINVALAIALAHISGLNLEU----- 102
 DB 800 GCCCAGATGGCG-----CTCTCAAAAGCGCTTGGCCACATCTCTGACACCCCTTGCTT 747
 QY 103 -----GLYThrGlySerValAlaMetValThrMetThrMetArgHisThr----- 117
 DB 746 GACTTTGGCAAGGATCAAGCTCTCAGAGCGATGTGGAGACAGTCACTCAATGATGCAT 687
 QY 118 ALAGLYGlnArgLeu-----HisAspLeuThrThrGlyLeuSerAlaAlaTrpVal 135
 DB 666 GGTGGGAAATCTCACTAGACCTGATTACTTTGAGGTAGAGACAGGCTGGGAGACC 627
 QY 136 -----AlaThrAsnGlyArgArgTrpArgThrGluArgGluMet 148
 DB 626 AGGTGAGCCAGAGCATGAGAGCGCGTGGAGGAGAGGCTTCTGTGACACATGGTTGCTC 567

QY 149 TyrGlyCysAspGlyTyrValArgAlaValGluIleThrHisGlyLyAsnGlyTrpHis 168
 DB 566 TGTGGGACCTTGGGATC-----TGGGGT 543
 QY 169 VALHISVALHISALALEUeMetPheSerGlyAspValSerGluAsnIleLeuGluSer 188
 DB 542 ACTAGGCTCCACAGATCTCTTGACCTCGGC----- 510
 QY 189 PheSerAspAlaMetPheAspArgTrpThrSerIlyLeuValSerLeuGlyPheAla 208
 DB 509 -----TCTCAGGCGCGGTGCTGGGCTCCACACTT 480
 QY 209 ProLeuArgAsnSerGlyLyLeuAspValArgValIleGlyGlyGluAlaAspGlnVal 228
 DB 479 CCGTCCCGCCCTCAGGACGAGACTGAGCTCA-----GACGAGATC 438
 QY 229 LeuAlaAlaTyrLeuThrThrIleAlaSerGlyVal-----GlyMetGluValIlySerGly 247
 DB 437 CTCTGAGGGGATCTTGGCCGGCTGTGACCAACATTAACAGTTGATCTTGTTCGCTT 378
 QY 248 AspGlyLySerGlyArgHisGlyAsnArgAlaProTrpGluIleAlaValAspAlaVal 267
 DB 377 GACGGGGATGGCAAAAGGGGAGACCGAGGTCACTGAGGGCTAGGCTCAGC----- 324
 QY 268 GlyGlyAspProGlnAlaLeuGluLeuTrpArgGluPheGluPheGlySerMetGlyArg 287
 DB 323 -----TGG-----AGG 318
 QY 288 ArgAlaIleAlaTrpSerArgGlyLeuArgAlaArgAlaGlyLeuGlyValLeuThr 307
 DB 317 GGTGGGCTCAGCTGGGGAACGTGGCTCAGGACGAGAACTGGGGCAGGA----- 270
 QY 308 AspAlaGlnIleValGluGlnGluGluSerAlaProValMetValAlaIleIleProAla 327
 DB 269 -----GTCCATTCAGGACGCGTAACT-----CCGGCG 240
 QY 328 ArgSerTrpMetMetIleArgThrCysAlaProTyrValPheGlyGluIleGluGlyLeu 347
 DB 239 CGC-----GAGCGGCGCGGGCTG 222
 QY 348 ValGluAlaGlyAlaThrTrp-GluAsnLeuArgAspHisIleHisTyrArgLeuProAl 367
 DB 221 GTCACTCCCGGCTTGGCTGGGGCGCGCGCTTGGCGGCTGCTGCTCGGTGAACGC 162
 QY 367 ALAAspValAlaArgProProIleIleSerValArgLyS 379
 DB 161 AGGAGCCAGAGGCGCCACAGGCTCGGGGTAGAGACG 125
 RESULT 10
 BZ554247/c
 LOCUS BZ554247 893 bp DNA linear GSS 17-DEC-2002
 DEFINITION pacsl-60_4590.y2 pacsl-60 Pseudomonas aeruginosa genomic clone
 pacsl-60_4590, genomic survey sequence.
 ACCESSION BZ554247
 VERSION BZ554247.1 GI:27161340
 KEYWORDS GSS.
 SOURCE Pseudomonas aeruginosa
 ORGANISM Pseudomonas aeruginosa
 Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
 Pseudomonadaceae; Pseudomonas.
 REFERENCE 1 (bases 1 to 893)
 Spencer, D.H., Raymond, C.K., Smith, E.E., Sims, E.E., Hastings, M.,
 Burns, J.L., Kaul, R. and Olsen, M.V.
 Whole-Genome-Sequence Variation among Multiple Isolates of
 Pseudomonas aeruginosa library
 J. Bacteriol. (2002) In press
 CONTACT: Chris K. Raymond
 Genome Center
 University of Washington
 Box 352145, Seattle, WA 98105-2145, USA
 Tel: 2062216954
 Fax: 2066857244
 Email: craymond@u.washington.edu

```

FEATURES
  source
    Class: shotgun.
    Location/Qualifiers
      1. 893
        /organism="Pseudomonas aeruginosa"
        /mol_type="genomic DNA"
        /strain="1-60"
        /db_xref="taxon:287"
        /clone="pacs1-60_4590"
        /clone_1ib="pacs1-60"
        /note="Clinical isolate 1-60 Whole genomic shotgun
        library."
BASE COUNT      151 a      332 c      270 g      138 t      2 others
ORIGIN
Alignment Scores:
  Pred. No.:      1.03      Length:      893
  Score:          118.00     Matches:    74
  Percent Similarity: 34.07%  Conservative: 19
  Best Local Similarity: 27.11%  Mismatches: 116
  Query Match:      5.89%      Indels:     66
  DB:               29         Gaps:       10
US-10-007-527a-2 (1-379) x BZ554247 (1-893)
QY      16 ProvalleuValserSerAspLySAArgLYtleArgHISgluleuArgProLySLeuGln 35
      730 CCAAGTCCCTGTCGCGGCAACCCCTTCGGGGCGTTCAATTCGGGAC-AGCCAGAAGCGCACG 672
QY      36 GlnlleHrHrser-----GlnThrpheasnaIaCysgIyArpProIle 50
      671 CAGGCGACCCCGACAGTCTCGGCAAGTCTCGGCTTCCAGCGCGGTGAACGGCCCGC 612
QY      51 SerGlyValasnGlyValThrIleValasnGlyProLySgIserGlyPheGlyGlyLeu 70
      611 GTTGGC-----GTAGNCGCGAGAGGCGCGGCGCGGAGCGGAGC 570
QY      71 ArgSerCysgIyLySgIyTrpIleCysProCysCysAlaGlyLySValGlyAlaHisArg 90
      569 CAGNCT-----GCCGTCGCGATTGAGAGCACCGATC 540
QY      91 AlaAspGlnIleSerGlnValAlaHisGlnLeuGlyHrGlySerValAlaMetVal 110
      539 GTTCAGGGATGATCGCCAGGTCTGAGACATCCAGCCGAGAGGCG---GTTGTAAGAGTC 483
QY      111 ThrMetThrMetArgHisThrAlaGlyGlnArgLeuHisAspLeuTrpHrGlyLeuSer 130
      482 GATACAGACCGCCCGCGCGCGCGCGCGTGAACAGCA-GTTGTCCGGAACAGCGCGCAT 424
QY      131 AlaAlaTrpLySAlaAlaThrAsnGlyArgArgTrpArgHrGlnArgGluMetTrpGly 150
      423 GCAGGTTGGCGCGCGGAGCGCAGGCGCTCGCATCCAGCGCGCGCATCTCCGCCAGGG 364
QY      151 CysAspGlyTrpValAlaGala---ValGlnlleHrHrHISgIyLySAsnGlyTrpHisVal 169
      363 CCGGGGCGAGCAGGCGCGCGCGCTGTTCGGTAGTCCGGGGGCA---GGTTGGCAC--- 310
QY      170 HisValHisAlaLeuMetPheSerClyAspValSerGlnAsnIleLeuGluSerPhe 189
      309 -----CTTGTTCACGCA 298
QY      190 SerAspAlaMetPheAspArgTrpTrpHrSerLySLeuValSerLeuGlyPheAla----- 207
      297 TCCAGGGCAGACACAGCGTCTGGGGCGTTCAAGAGATCGCCCGCGGTGGCGCGGTGA 238
QY      208 -----AlaProLeuArgAsnSerGlyGlyLeuAspValArgLySAlleGly 222
      237 GGTGGCGGAGAGATCGCCCACTTCTCGCAGTGTGGCGCTTCGGCTCGCGCTCTGAGC 178
Db      223 GlyGlnAlaAspGlnValLeuAlaAlaTrpLeuThrLySAlleAlaSerGlyValGlyMet 242
      177 GGGCGGCGAGCGCGGTTGCGACAGCG-----CGGCTTGCTT 139
QY      243 GlnValGlySerGlyAspGlyLySserGlyArgHisGlyAsnArgAlaProTrpGlnIle 262

```

```

Db      138 CGAGACGGGCGAGGCGCTTCGCCGTCGGGTCGCGAGACGTAAGGCGACCG----- 88
QY      263 AlaValaAspAlaValaGlyGlyAspProGlnAlaLeuGln 275
      87 -----GGGATTCACACTAGTTCTTAGAG 67
RESULT 11
BQ936859/c 1306 bp mRNA linear EST 21-AUG-2002
LOCUS
DEFINITION
  AGENCOURT 10015321 NCI CGAP Mam2 Mus musculus cDNA clone
  IMAGE:6487576 5', mRNA sequence.
ACCESSION
  BQ936859
VERSION
  BQ936859.1 GI:22352242
KEYWORDS
  EST.
SOURCE
  Mus musculus (house mouse)
  Mus musculus
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
  1 (bases 1 to 1306)
  NIH-MGC http://mgs.nci.nih.gov/.
  National Institutes of Health, Mammalian Gene Collection (MGC)
  Unpublished
  CONTACT: Robert Strausberg, Ph.D.
  Email: cgabs-remail.nih.gov
  Tissue Procurement: Gilbert Smith, Ph.D.
  cDNA Library Preparation: Life Technologies, Inc.
  CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
  DNA Sequencing by: Agencourt Bioscience Corporation
  Clone distribution: MGC clone distribution information can be
  found through the I.M.A.G.E. Consortium/LLNL at:
  http://image.llnl.gov
  Plate: LLNL4032 row: m column: 17
  High quality sequence stop: 125.
FEATURES
  source
    1. 1306
      /organism="Mus musculus"
      /mol_type="mRNA"
      /strain="FVB/N-3"
      /db_xref="taxon:10090"
      /clone="IMAGE:6487576"
      /tissue_type="tumor, biopsy sample"
      /dev_stage="5 months"
      /lab_host="DH10B"
      /clone_1ib="NCI CGAP Mam2"
      /note="Organ: mammary; Vector: pCMV-SPORT6; Site 1: SalI;
      Site 2: NotI; Cloned unidirectionally. Primer: Oligo dT.
      Library constructed by life Technologies. Investigator
      providing samples: Gilbert Smith, NIH"
BASE COUNT      216 a      571 c      327 g      190 t      2 others
ORIGIN
Alignment Scores:
  Pred. No.:      2.1      Length:      1306
  Score:          117.00     Matches:    91
  Percent Similarity: 32.38%  Conservative: 22
  Best Local Similarity: 26.07%  Mismatches: 100
  Query Match:      5.84%      Indels:     137
  DB:               13         Gaps:       19
US-10-007-527a-2 (1-379) x BQ936859 (1-1306)
QY      45 AlaCysGlyArpProIleSerGlyValasnGlyValThrIleValasnGlyPro----- 62
      947 GCATGTGGCGCA-----GCGGGGTTCTCGTGTGCGGTCACATGACGGGCGCGCA 894
Db      63 -----LySgIySerGlyPheGlyGly-----LeuArgSerCys 73
      893 TGGCGCGGAGACGATGGGATGAGGTGATACCGCGCGGAAAGGATTCGACAAAGCC 834
QY      74 GlyLySgIyTrpIleCysProCysCysAlaGlyLySValGlyAlaHisArgAlaAspGly 93
      833 GGG-----GTGGTTACGGCGGACGGAT--- 810

```

RESULT	12
BF783863/c	
LOCUS	BF783863
DEFINITION	Bf783863 602109787g1 NCI_CGAP Kid14 Mus musculus cDNA clone IMAGE:4238036 5'', mRNA sequence.
ACCESSION	BF783863
VERSION	BF783863.1 GI:12088899
KEYWORDS	EST,
SOURCE	Mus musculus (house mouse)
ORGANISM	Mus musculus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.	

Db 1051 AGGGGGAGTTCCTGCGCGCTTGGGAGG-----GTGGCGTGGAG-----GTGGGTGCT 1004

QY		158	ValGluIleThrHisSgLYlyAsnGLYTThPHisValHisValHisAlaLeuMetPhe	177
Db		1003	 ::: GTGGCTGTGTGGAGTGTGAAGCGTGTTGGTGTGT-GTG-----	963
QY		178	SerGIysAspValSerGIuAsnIleLeuJusSerPheSerAspAlaMetPheAspArgTrp	197
Db		963	-----	963
QY		198	ThrSerLysLeuValSerLeuGLyPheAlaAlaProLeuAArgAsnSerGLyILeuAsp	217
Db		962	 -----TTGGTGGGGGTGTGGGGGGCGTCGC GTGTGGCGAGAAGGTGGGGGTGT---	915
QY		218	ValArgLysIleGLyGLyGLuAlaAspGlnValLeu-----AlaAlaTYrLeuThr	234
Db		914	 ::: -----GGGGGTGGCGCGCCGTGTGTGTGTGTGGCGCGGTGTGGGGTGGGGGTG	864
QY		235	LysIleAlaSerGLyValGLyMetCGLuValGLySer-----GLyAspGLy	249
Db		863	 CGGGTCCGGGTGTGCCCGGGGGGGAGAGGTCCGGGTCGGGTGGGGGGCCCTGGGGGGGGGG	804
QY		250	LysSerGLyAArgHisGLyAsnAArgAlaProTrpGluIleAlaVal-AspAlaValGLyGI	269
Db		803	 ::: AGGTCCGGGGCGCGGGGGGTGTGGCGCGCGGTG-----GTGCCGGGGCGTGTGGGGGG	753
QY		269	YAspProGlnAlaLeuGluLeuTrp-----Ar	278
Db		752	 CGGGAG-----GAGAGGTGTGTGGCGCGGGGTGTGGGGCGCCGTGTGAAGGGGGCG	702
QY		278	GGLIPheGluPheGLySerMetGLyAArgAArgAlaIleAlaTrpSer-----	293
Db		701	 TGAGGGCGCGGGGGGTGTGTGTGGGGCGGACCGCGGGGGCGTGTGGCGCCGGGGCGCGGGGA	642
QY		294	-----ArgGLyLeuAArgAlaAArgAl	300
Db		641	CGGGGGTG	582
QY		300	aGLyLeuGLy	303
Db		581	 CGGGCGGTGGG	572
RESULT 13				
LOCUS	BUT05875/c			
DEFINITION	BU705875	689 bp	mRNA	linear EST 09-OCT-2002
ACCESSION	UI-M-FR0-cai-a-04-0-U.r1 NIH_BMAP_FR0	Mus musculus	cDNA clone	
VERSION	BU705875			IMAGE: 6412299 5', mRNA sequence.
KEYWORDS	BU705875.1 GI:23635790			
SOURCE	EST.			
ORGANISM	Mus musculus (house mouse)			
TITLE	Mus musculus			
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
JOURNAL	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
COMMENT	NIH-MGC http://mgc.mci.nih.gov/ National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished Contact: Robert Strausberg, Ph.D. Email: rsra@bbs-rcmail.nih.gov Tissue Procurement: Dr. Jim Lin, University of Iowa cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa DNA Sequencing by: Dr. M. Bento Soares, University of Iowa Clone Distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LML at: http://image.llnl.gov This clone was contributed by the Brain Molecular Anatomy Project (BMAP)			
FEATURES	Seq primer: PYX-5.			
source	Location/Qualifiers			
	1..689			
	/organism="Mus musculus"			
	/mol_type="mRNA"			
	/strain="CS7BL/6"			

Db	336	CGCTTGAGGGGGGATGGCGAAGCGGGGAGCCAGAGGCTGACGTGGAGGGCTAGGGTACG-	278
Qy	246	ecGlyAsnGlyysSerGlyArgHisGlyAsnArgAlaProTrrpGluIleAlaValAspa	266
Db	396	GAGTCCTCGAGAGGGACATCTTGCCCGGCTGTTCACCAATAACAGTTGACTCTTTGTT	337
Qy	227	InValIleuAlaAlaTyrIleuThrLysIleAlaSerGlyVal--GlyMetGluValGlys	246
Db	438	AACTTCGCGCCCGCTCAGCAGAGACTGGAGCTCAGA--GAGC	397
Qy	207	IaAlaProLeuArgAsnSerGlyGlyLeuAspValArgLysIleGlyGlyAlaAspG	227
Db	501	GGGGTACTAGGCTCCACAGACATCTCTTGACTCGGGC--	464
Qy	167	IuSerPheSerAspAlaMetPheAspArgTrrPheSerLysIleuValSerIleuGlyPheA	207
Db	463	-----TTCAGGGGCGCGTCCGTGGCTCCC	439
Qy	207	IaAlaProLeuArgAsnSerGlyGlyLeuAspValArgLysIleGlyGlyAlaAspG	227
Db	438	AACTTCGCGCCCGCTCAGCAGAGACTGGAGCTCAGA--GAGC	397
Qy	227	InValIleuAlaAlaTyrIleuThrLysIleAlaSerGlyVal--GlyMetGluValGlys	246
Db	396	GAGTCCTCGAGAGGGACATCTTGCCCGGCTGTTCACCAATAACAGTTGACTCTTTGTT	337
Qy	246	ecGlyAsnGlyysSerGlyArgHisGlyAsnArgAlaProTrrpGluIleAlaValAspa	266
Db	336	CGCTTGAGGGGGGATGGCGAAGCGGGGAGCCAGAGGCTGACGTGGAGGGCTAGGGTACG-	278

QY 266 laValglYgIaSPProGlnAlaLeuGluLeuTrpArgGluPheGluPheGlySerMetG 286
 Db 277 -----TGG----- 275
 QY 286 lYArGaRgaLaIleAlaTrpSerArgGlyLeuArGaLaArGaLaGlyLeuGlu 306
 Db 274 --AGGGCTGGCTCAGCTGGGGAACCTGGCTCAGGAGAACTGGGGCAGGA----- 224
 QY 306 eUThraSPAlaGlnIleValGluGlnGluSerAlaProValMetValAlaIleLeu 326
 Db 223 -----GTCCAAATCCAGGAGCCGTAACT-----C 199
 QY 326 roAlaRgSerTrpMetMetIleArgThrCysAlaProTyrValPheGlyGluIleLeu 346
 Db 198 CGGGCGGC-----GAGCGGCGCG 181
 QY 346 lYLeuValGluAlaGlyAlaThrTrp-GluAsnLeuArgAspHisLeuIstYrArgLeu 365
 Db 180 GGCTGGTCACTCCCGGGTTGGCTGGGGCGCGGCGCTTGGCGCTGCTCCGGGTG 121
 QY 366 ProAlaAlaSPValArgProProIleIleSerValArgLys 379
 Db 120 AACGAGAGCCACAGGCGCACGAGCTCGGGGTGAGACAG 79
 RESULT 14
 B1411296/c 867 bp mRNA linear EST 14-AUG-2001
 LOCUS 602964678F1 NCI_CGAP_Lu33 Mus musculus cDNA clone IMAGE:5120034 5',
 DEFINITION mRNA sequence.
 ACCESSION B1411296
 VERSION B1411296.1 GI:15172219
 KEYWORDS EST.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Bularyca; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 NIH-MGC http://mgc.nci.nih.gov/
 1 (bases 1 to 867)
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished
 Contact: Robert Strausberg, Ph.D.
 Email: cgabs-remail.nih.gov
 Tissue Procurement: Gilbert Smith, Ph.D.
 cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
 Bonaldo, Ph.D.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNL at:
 http://image.jnl.gov
 Plate: LHAM11292 row: 1 column: 19
 High quality sequence start: 32
 High quality sequence stop: 861.
 Location/Qualifiers
 1..867
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="CZECH 11"
 /db_xref="taxon:10090"
 /clone="IMAGE:5120034"
 /tissue_type="pooled lung tumors"
 /lab_host="DH10B (phage-resistant)"
 /clone_lib="NCI CGAP Lu33"
 /note="Organ: lung; Vector: pT73D-Pac (Pharmacia) with a
 modified polylinker; Site 1: NotI; Site 2: EcoRI; 1st
 strand cDNA was prepared from mRNA obtained from pooled
 lung tumors with a Not I - oligo(dT) primer [5'
 TGTACCAATCTGAGAGTGGAGCGCGCTGTGTTTCTTTTCTTTT 3'].
 Double-stranded cDNA was ligated to Eco RI adaptors
 (Pharmacia), digested with Not I and cloned into the Not
 I and Eco RI sites of the modified pT73 vector. Library
 went through one round of normalization, and was

BASE COUNT 171 a 292 c 234 g 170 t
 ORIGIN
 Alignment Scores:
 Pred. No.: 1.51 Length: 867
 Score: 116.00 Matches: 83
 Percent Similarity: 32.31% Conservative: 22
 Best Local Similarity: 25.54% Mismatches: 89
 Query Match: 5.79% Indels: 132
 DB: 12 Gaps: 18
 US-10-007-527a-2 (1-379) x B1411296 (1-867)
 QY 69 GlyLeuArgSer-----CysGly-----LysGlyTrpIleCysProCys 81
 Db 710 GGCAAAAGGTCACTCAGCTCAGAGCAGTGTGGAGCATGCTCAAAATG--TGTCCATGG 654
 QY 82 CysAlaGlyLysValGlyAlaHisArgAlaAspGluIleSerGlnValAlaHisGln 101
 Db 653 TGC----- 651
 QY 102 LeuGlyThrGlySerValAlaMetValThrMetThrMetArgHisThrAlaGlyGlnArg 121
 Db 650 ---GGAACATCTCACTACACCTGATCTGTAGAGGTA-----GGACAAAGC 606
 QY 122 LeuHisAspLeuTrpThrGlyLeuSerAlaAlaTrpLysAlaAlaThrAsnGlyArgArg 141
 Db 605 CTGGAGAGACCA-----GGCTGAGCCAGAGCATGTAGAG-----CGC 571
 QY 142 Trp-ArgThrGlu-----ArgGluMetCysGly-TyrValArgAlaValG 159
 Db 570 TGGTACGAGAGAGCTTCTTCAACATGTTGCTCTGT----- 533
 QY 159 IuIleThrHisGlyLysAsnGly---TrpHisValHisValHisAlaLeuMetPheS 178
 Db 532 -----GGGCACTTGGCATTTGGGGTACAGCTCCACAGCATTTCTTGGACT 484
 QY 178 eRgLYaSPValSerGluAsnIleLeuGluSerPheSerAspAlaMetPheAspArgTrpT 198
 Db 483 CGGGC----- 479
 QY 198 HisSerLysLeuValSerLeuGlyPheAlaAlaProLeuArgAsnSerGlyLysLeuAspY 218
 Db 478 --TTCAGAGCCCGTCCGCTCCCACTTCCGCCCGCTCAGGAGCATCTGGAG 421
 QY 218 aArgTrpIleGlyGlyGluAlaAspGlnValIleuAlaAlaTyrLeuThrLysIleAlaS 238
 Db 420 TCAGA-----GACGAGTCTCTGAGGGAGATCTGCCCGGCTTTGA 379
 QY 238 eRgLYVal---GlyMetGluValGlySerGlyAspGlyLysSerGlyArgHisGlyAsnA 257
 Db 378 CCAACATAAACAGTTGTACTCTTGTTCCTTGAGCGGGAGATGGGCAACGGGGAGCCA 319
 QY 257 rGAlaProTrpGluIleAlaValaAspAlaValGlyLysProGlnAlaLeuGluLeuT 277
 Db 318 GGGTCACTGAGAGGCTAGGTCAGC-----T 292
 QY 277 rPArgGluPheGluPheGlySerMetGlyArgArgAlaIleAlaTrpSerArgGlyLeuA 297
 Db 291 GG-----AGGGCTGGCTCAGCTGGGAACTGGAGCTCA 259
 QY 297 rGAlaArgAlaGlyLeuGlyAlaGluLeuThrAspAlaGlnIleValGluGlnGluS 317
 Db 258 GGCAGAGAACTGGGGCAGGA-----GTCCAAATCCAGGACAGC 223
 QY 317 eRAlaProValMetValAlaIleIleProAlaArgSerTrpMetMetIleArgThrCysA 337
 Db 222 CGTAACCT-----CCGGCGGC----- 206
 QY 337 lAProTrValPheGlyGluIleLeuGlyLeuValGluAlaGlyAlaThrTrp-GluAsn 356
 Db 205 -----GAGCGGCGGGCTGTGTACTCCGGGTTCCGCTGGGGCCGG 163

```

QY      357 LeuArgAspHisIleuHisIstYrArgLeuProAlaAlaAspValArgProIleIleSer 376
Db      162 GGCAGCTTGCGGCTGCTGCTCGGAGTGAAGCAGAGCCACGAGGCCACCGAGGCTCGG 103
QY      377 ValArgLys 379
Db      102 GTGAGACAG 94

RESULT 15
BG684181/c 1199 bp mRNA linear EST 01-MAY-2001
LOCUS 602635811F1 NIH_MGC_46 Homo sapiens cDNA clone IMAGE:4763523 5'
DEFINITION mRNA sequence.
ACCESSION BG684181
VERSION BG684181.1 GI:13915578
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
JOURNAL NIH-MGC http://mgi.nci.nih.gov/
COMMENT National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished
Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: Louis M. Staudt, M.D., Ph.D.
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LIML)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LIML at:
http://image.llnl.gov
plate: LCM1619 row: f column: 04
High quality sequence start: 51
High quality sequence stop: 147.
Location/Qualifiers
1. 1199
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:4763523"
/tissue_type="primary B-cells from tonsils (cell line)"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_48"
/note="Organ: B-cells; Vector: pOTB7; Site: 1: XhoI;
Site 2: EcoRI; cDNA made by oligo-dT priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCAAGAG(G). Size-selected >500bp
for average insert size 1.8kb. Library constructed by Ling
Hong in the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies).
Note: this is a NIH_MGC Library."
BASE COUNT 263 a 502 c 300 g 134 t
ORIGIN

Alignment Scores:
Pred. No.: 2.32 Length: 1199
Score: 116.00 Matches: 75
Percent Similarity: 31.53% Conservative: 24
Best Local Similarity: 23.89% Mismatches: 111
Query Match: 5.79% Indels: 104
DB: 10 Gaps: 17

US-10-007-527a-2 (1-379) x BG684181 (1-1199)
QY      47 GLYArgProIleSerGlyValAsnGlyValThrIleValAsnGlyProLys---GlySer 65
Db      1074 GGGCGGCGCGGGTGGCGGGTGGGGGGCGCGTGTGTGCGGGCGGCGCGCGTGTGGCTGC 1015
QY      66 GlyPheGlyGly-----LeuArgSerCysGlyLysGlyTyrIleCys 79

```

```

Db      1014 GGGTGGGGGGGCTGCGCGTGGCTGGCGCGCGGCGCGCGCGCTGGGGGCGCGTGT 955
QY      80 ProCysCysAlaGlyLysValGlyValAlaHisArgAlaAspGluIleSerGluValAla 99
Db      954 TTGCAGTGTTCGGGCGCGGTGGGGTGGGCGCGCGGAGTGTTCGGGTGGCGGCG 895
QY      100 -----HisGluLeuGlyThrGlySerVal 107
Db      894 GCGTGTGGCGTCCGCTTGGCGAAGCGGAGCGCTGGGGCGCTGAGCGCTTGGCGGT--- 838
QY      108 AlaMetValThrMetThrMetArgHisThrAlaGlyGlnArgLeuHisAspLeuTyrThr 127
Db      837 GCGTGTGTGGCACTCGGGGCTTAGG----- 814
QY      128 GlyLeuSerAlaAlaTyrLysAlaAlaThrAsnGlyArgArgTyrArgThrGluArgGlu 147
Db      813 GGTGCTCTCGACGTGGGGCGATGTCTGGCGGGGCGCTGTCTGGGGGCC-----CGA 760
QY      148 MetTyrGlyCysAspGlyTyrValArgAlaValGluIleThrHisGlyLysAsnGlyTyr 167
Db      759 TTGTGTGGGGCGCTAAGTGC-----ACTCTGGACTTGGCTGTCTGG 718
QY      168 HisValHisValHisAlaLeuLeuMetPheSerGlyAspValSerGluAsnIleLeuGlu 187
Db      717 -----CATGCGTGTGGCGCGCGCTGGCGGGCTGCGCC-----CAG 685
QY      188 SerPheSerAspAlaMetPheAspArgTyrThrSerLysLeuValSerLeuGlyPheAla 207
Db      684 GGCACCTGTGGGTGGGTGGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 625
QY      208 AlaProLeuArgAsnSerGlyLysLeuAspValArgLysIleGlyGlyGluAlaAspGln 227
Db      624 TGTCCG----- 619
QY      228 ValLeuAlaAlaTyrLeuThrLysIleAlaSerGlyValGlyMetGluValGlySerGly 247
Db      618 -----GCTGGCATGTCTGGGGAGCGCGGT 595
QY      248 -----AspGlyLysSerGlyArgHisGlyAsn-----ArgAlaProTyr 260
Db      594 CTTGTGTATGTGGTGGGAGGTTGGCATGTGCACGTGAGCGCATGTGATGCGGTGGGTGG 535
QY      261 GluIleAlaValAspAlaValGlyLysAspProGlnAlaLeuGluLeuTyrArgGluPhe 280
Db      534 TGTCTTACAGAGAGACGGGTGACAGGTGACACGACGCTA----- 493
QY      281 GluPheGlySerMetGlyArg-----ArgAlaIleAlaTyr-----SerArgGlyLeu 296
Db      492 -----GCCGATGCAGAGACGGTGGCTGTGCTGGGTGTGTGTGACACCGTGTGGAGGCTC 439
QY      297 ArgAlaArgAlaGlyLeuGlyValGluIleuThrAspAlaGlnIleValGluGlnGluGlu 316
Db      438 GAGGGCGCGGCTTGTCTGGGTAGG-----GGCAGCTTGTGTGTGTGTGTGTGTGTGTG 391
QY      317 SerAlaPro-----ValMetValAlaIleIleProAlaArg 328
Db      390 TCCGCCAAGAGACGTTGCTTACGTGAGTGTTCACACAGCGC 349

Search completed: August 22, 2003, 12:09:11
Job time : 1932 secs

```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 8, 2003, 21:40:18 ; Search time 60 Seconds

(without alignments)
1002.623 Million cell updates/sec

Title: US-10-007-527a-2

Perfect score: 2005
1 MTSVSAEHLSGKDRPPVLVS.....HHYRLPAADVPPILTSVRK 379

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-Processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_19Jun03:*

1: /SIDSI/gcgdata/geneseq-emb1/AA1980.DAT:*
2: /SIDSI/gcgdata/geneseq-emb1/AA1981.DAT:*
3: /SIDSI/gcgdata/geneseq-emb1/AA1982.DAT:*
4: /SIDSI/gcgdata/geneseq-emb1/AA1983.DAT:*
5: /SIDSI/gcgdata/geneseq-emb1/AA1984.DAT:*
6: /SIDSI/gcgdata/geneseq-emb1/AA1985.DAT:*
7: /SIDSI/gcgdata/geneseq-emb1/AA1986.DAT:*
8: /SIDSI/gcgdata/geneseq-emb1/AA1987.DAT:*
9: /SIDSI/gcgdata/geneseq-emb1/AA1988.DAT:*
10: /SIDSI/gcgdata/geneseq-emb1/AA1989.DAT:*
11: /SIDSI/gcgdata/geneseq-emb1/AA1990.DAT:*
12: /SIDSI/gcgdata/geneseq-emb1/AA1991.DAT:*
13: /SIDSI/gcgdata/geneseq-emb1/AA1992.DAT:*
14: /SIDSI/gcgdata/geneseq-emb1/AA1993.DAT:*
15: /SIDSI/gcgdata/geneseq-emb1/AA1994.DAT:*
16: /SIDSI/gcgdata/geneseq-emb1/AA1995.DAT:*
17: /SIDSI/gcgdata/geneseq-emb1/AA1996.DAT:*
18: /SIDSI/gcgdata/geneseq-emb1/AA1997.DAT:*
19: /SIDSI/gcgdata/geneseq-emb1/AA1998.DAT:*
20: /SIDSI/gcgdata/geneseq-emb1/AA1999.DAT:*
21: /SIDSI/gcgdata/geneseq-emb1/AA2000.DAT:*
22: /SIDSI/gcgdata/geneseq-emb1/AA2001.DAT:*
23: /SIDSI/gcgdata/geneseq-emb1/AA2002.DAT:*
24: /SIDSI/gcgdata/geneseq-emb1/AA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2005	100.0	379	23	ABB84278 Rhodococcus AN12 r
2	609.5	30.4	459	23	ABB84280 A. pyogenes pAPI r
3	399.5	19.9	494	19	AAW73071 B. brevis essential
4	326	16.3	427	21	AA9797536 B. lactofermentum
5	326	16.3	427	23	ABB06341 Brevibacterium lac
6	326	16.3	427	23	AA813530 S. nigrofasciatum
7	322	16.1	451	23	ABB84283 S. nigrofasciatum
8	318	15.9	427	21	AA9797537 B. lactofermentum
9	318	15.9	427	23	ABB06342 Brevibacterium lac

10	318	15.9	427	23	AAE13531
11	316	15.8	456	23	ABB84281
12	282	14.1	466	23	AAU11039
13	266.5	13.3	528	23	ABB84282
14	134	6.7	314	14	AA931143
15	103	5.1	647	24	AAU56572
16	100.5	5.0	577	17	AAU05554
17	100	5.0	1568	22	ABG03109
18	97	4.8	713	21	AAV58582
19	95.5	4.8	3798	21	AAV58577
20	94.5	4.7	431	23	ABP65883
21	94.5	4.7	1721	19	AAW52847
22	94	4.7	386	24	AAE30185
23	93.5	4.7	356	22	ABE10262
24	93.5	4.7	356	23	ABP65849
25	93.5	4.7	1049	24	ABR41258
26	92.5	4.6	579	20	AAV35500
27	92	4.6	740	22	ABG24479
28	91.5	4.6	444	23	ABP65024
29	91.5	4.6	527	22	AAU59539
30	91.5	4.6	622	22	AAU71880
31	91	4.5	956	22	AAU36332
32	90.5	4.5	415	22	AAU41823
33	90.5	4.5	729	22	ABE59500
34	90	4.5	668	22	AAU63901
35	89.5	4.5	361	21	AAU40780
36	89.5	4.5	369	21	AAU40779
37	89.5	4.5	375	21	AAU40778
38	89.5	4.5	509	22	AAU67532
39	89.5	4.5	585	22	AAU92278
40	89	4.4	350	22	AAU61038
41	89	4.4	898	23	AAU09878
42	89	4.4	3562	22	AAU82213
43	89	4.4	3739	21	AAU18638
44	89	4.4	3739	21	AAU77193
45	89	4.4	3739	21	AAU77201

ALIGNMENTS

RESULT 1	ABB84278	standard; Protein; 379 AA.
ID	ABB84278	
XX	ABB84278;	
AC	13-JAN-2003	(first entry)
DT	13-JAN-2003	
DE	Rhodococcus AN12 replication protein Rep.	
XX		
KW	Plasmid stability protein; replication protein; ethylene forming enzyme;	
KW	carotenoid biosynthesis enzyme; isoprenoid; pyruvate decarboxylase;	
KW	polynitroxylalanoic acid synthase; PNA synthase; nitrile hydratase;	
KW	alcohol dehydrogenase; terpene synthase; cholesterol oxidase;	
KW	shuttle vector; Rep protein.	
OS	Rhodococcus erythropolis.	
XX		
XX	W0200255709-A2.	
PD	18-JUL-2002.	
XX		
PF	12-DEC-2001; 2001MO-US47868.	
XX		
PR	12-DEC-2000; 2000US-254868P.	
XX		
PA	(DUPO) DU POINT DE MEMOIRS & CO E I.	
XX		
PI	Bramucci MG, Cheng Q, Kostichka KN, Tomb J;	
XX	WPI, 2002-557827/59.	
DR	N-PSDB; ABQ76122.	

B. lactofermentum
S. lividans pji101
Replication protei
S. phaeochromogen
Sequence encoded b
Lung cancer associ
R. capsulatus acet
Novel human diagno
Sorangium cellulos
Bifidobacterium lo
A. mediterranei ri
Human PHD1 protein
Human CDNA SEQ ID
Human polyprotein
Human DITP intrac
Chlamydia pneumoni
Novel human diagno
Bifidobacterium lo
Propionibacterium
Human PEPCK-cytoso
Pseudomonas aerugi
Propionibacterium
Drosophila melanog
Propionibacterium
Zea mays protein f
Zea mays protein f
Propionibacterium
C glutamicum prote
Propionibacterium
Novel human secret
Polyketide synthas
Amino acid sequenc
S. venezuelae plik
S. venezuelae plik

XX New nucleic acid molecule encoding replication protein/plasmid
PT stability protein, useful in cloning and expression vectors,
PT particularly shuttle vectors for expression of heterologous genes in
PT Rhodococcus species -
XX
XX Claim 4; Page 64-65; 96pp; English.
XX
XX This invention describes a novel nucleic acid encoding a replication
CC protein or a plasmid stability protein. The product of the invention is
CC useful for expression of nucleic acid such as genes encoding enzymes
CC involved in the production of isoprenoid molecules, polyhydroxyalkanoic
CC acid (PHA) synthases, carotenoid biosynthesis enzymes, nitrile
CC hydratases, ethylene forming enzyme, pyruvate decarboxylase, alcohol
CC dehydrogenase, terpene synthases, and cholesterol oxidase in an
CC Actinomycetales bacteria. The replication protein or plasmid stability
CC protein are useful in cloning and expression vectors and particularly in
CC shuttle vectors for the expression of homologous and heterologous genes
CC in Rhodococcus sp. This sequence represents the Rhodococcus AN12 Rep
CC protein described in the disclosure of the invention.
XX
XX
SQ Sequence 379 AA;
Query Match 100.0%; Score 2005; DB 23; Length 379;
Best Local Similarity 100.0%; Pred. No. 1.5e-190;
Matches 379; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MTSVSAEHLGSKRPPVLYSSDRGRHRELPKLOTTSETFNACGRPISGVNGYTVN 60
DB 1 MTSVSAEHLGSKRPPVLYSSDRGRHRELPKLOTTSETFNACGRPISGVNGYTVN 60
QY 61 GPKSGGFGGLRSCGKGMICPCCAKGVAGRADEISQVVAHQLTGSGVAWVTMTMTAAG 120
DB 61 GPKSGGFGGLRSCGKGMICPCCAKGVAGRADEISQVVAHQLTGSGVAWVTMTMTAAG 120
QY 121 RLHDLMTGLSAAMKAATNGRRRTERTERMVGGDGYVRAVEITHGKNGHVAHLLMFSGD 180
DB 121 RLHDLMTGLSAAMKAATNGRRRTERTERMVGGDGYVRAVEITHGKNGHVAHLLMFSGD 180
QY 121 VSHNIIIESFSDAMPDRRTKSLVGLGFAAPLRNSGGIDVRKIGEAQOVLAAYITKIASGV 240
DB 181 VSHNIIIESFSDAMPDRRTKSLVGLGFAAPLRNSGGIDVRKIGEAQOVLAAYITKIASGV 240
QY 241 GMEVSGDGKSGRGNRAPMEIIVDAVAGDPQALLETREPEFGSMGRRAIAMSRLGARA 300
DB 241 GMEVSGDGKSGRGNRAPMEIIVDAVAGDPQALLETREPEFGSMGRRAIAMSRLGARA 300
QY 301 GLGAEILTDAQIVQESAPVWVAIIPARSWMTIRTCAPYFGEILGLVEAGATWENLRDH 360
DB 301 GLGAEILTDAQIVQESAPVWVAIIPARSWMTIRTCAPYFGEILGLVEAGATWENLRDH 360
QY 361 LHYRLPADVDRPPIISVRK 379
DB 361 LHYRLPADVDRPPIISVRK 379
RESULT 2
ABB84280
ID ABB84280 standard; Protein; 459 AA.
XX
XX ABB84280;
XX
XX 13-JAN-2003 (first entry)
XX
DE A. pyogenes pAP1 replication protein.
XX
XX Plasmid stability protein; replication protein; ethylene forming enzyme;
KM carotenoid biosynthesis enzyme; isoprenoid; pyruvate decarboxylase;
KM polyhydroxyalkanoic acid synthase; PHA synthase; nitrile hydratase;
KM alcohol dehydrogenase; terpene synthase; cholesterol oxidase;
KM shuttle vector.
XX
XX Arcanobacterium pyogenes.
OS

XX WO200255709-A2.
XX
XX 18-JUL-2002.
XX
XX 12-DEC-2001; 2001WO-US47868.
XX
XX 12-DEC-2000; 2000US-254868P.
XX
XX (DUPLO) DU PONT DE NEMOURS & CO E. I.
XX
XX Brannucci MG, Cheng Q, Kostichka KN, Tomb J;
XX WPI; 2002-557827/59.
XX
XX
XX New nucleic acid molecule encoding replication protein/plasmid
PT stability protein, useful in cloning and expression vectors,
PT particularly shuttle vectors for expression of heterologous genes in
PT Rhodococcus species -
XX
XX
XX Example 5; Fig 4A; 96pp; English.

CC This invention describes a novel nucleic acid encoding a replication
CC protein or a plasmid stability protein. The product of the invention is
CC useful for expression of nucleic acid such as genes encoding enzymes
CC involved in the production of isoprenoid molecules, polyhydroxyalkanoic
CC acid (PHA) synthases, carotenoid biosynthesis enzymes, nitrile
CC hydratases, ethylene forming enzyme, pyruvate decarboxylase, alcohol
CC dehydrogenase, terpene synthases, and cholesterol oxidase in an
CC Actinomycetales bacteria. The replication protein or plasmid stability
CC protein are useful in cloning and expression vectors and particularly in
CC shuttle vectors for the expression of homologous and heterologous genes
CC in Rhodococcus sp. This sequence represents A. pyogenes pAP1 replication
CC protein described in the disclosure of the invention.
XX
XX
SQ Sequence 459 AA;
Query Match 30.4%; Score 609.5; DB 23; Length 459;
Best Local Similarity 36.3%; Pred. No. 1.1e-51;
Matches 148; Conservative 62; Mismatches 133; Indels 65; Gaps 15;

QY 5 SAHELSGKRRPPVLYSSD-----KRGIRHEIRPKLOTTSETFNACGR-PISGVNGY 56
DB 67 SAHEPLGNTVLTFFPVSNBSKKTAKSRSRERYELRDIAISTESVRSKGRVVAELVSL 126
QY 57 TIYNGPKSGGFGGLRSCGKGMICPCCAKGVAGRADEISQVVAHQLTGSGVAWVTMTMT 115
DB 127 RAKSDGKAGYGGELHTCGSVMACPVCSAKIARRKTDLOQVVDHAYKGMTVSMILITQR 186
QY 116 HTAGORLHDLMTGLSAAMKAATNGRRRTERTERMVGGDGYVRAVEITHGKNGHVAHLL 175
DB 187 HHKGGGLKHLMDALSTAMNRVTSGRRWIEFKQFGLVGVYRANEITHGKNGHVAHSHVLI 246
QY 176 MBSGDVSENILESF-----SPAMDRMTSKVSLGFAAPLRNSGG 215
DB 247 ISEKID---PLTSTFTVYQRRQGRRLPYPEIYNSSDPIAEKRWAGLAKGVDF-LBDSGG 302
QY 216 L-----DYRKIGEAQOVLAAYITKI-----ASGVAGVSGDGKSGRGNRAPMEIIVDA 266
DB 303 LDWTVAKDRAIYN-----YVSKMGSTDAIISSEVTLGSKKARANKRTPFOLLADI 354
QY 267 VG-GDPQALLETREPEFGSMGRRAIAMSRLGARAAGLGAELTDAQIVQESAPVWVAI 335
DB 355 LSLGDVDDDKLWKEYEKASFGRRALTWKSLRDMANLGVQSDDEIA-SPEIDDEAIALF 413
QY 326 PARSWMTIRTCAPYFGE--EILGLVAGA-----TWNLRDHLYLTP 366
DB 414 THDAMPQVRR-----FGAAELLDVTBSGGRAAAYRWLDFRE-IDWSLIP 455
RESULT 3
AAW73071
ID AAW73071 standard; Protein; 494 AA.

```
XX
AC AAW73071;
XX
DT 05-JAN-1999 (first entry)
XX
DE B. breve essential region gene protein.
XX
KM Essential region gene; shuttle vector.
XX
OS Bifidobacterium breve.
XX
PN JP10262670-A.
XX
PD 06-OCT-1998.
XX
PF 27-MAR-1997; 97JP-0091387.
XX
PR 27-MAR-1997; 97JP-0091387.
XX
PA (HONS ) YAKULT HONSHA KK.
XX
DR WPI; 1998-587288/50.
XX
N-PSDB; AAV58945.
XX
PS Shuttle vector for a Bifidobacterium species - contains specific
XX elements from existing plasmids e.g. pNBbl of ATCC 15698
XX
CC Claim 5; Page 14-15; 17pp; Japanese.
XX
CC This sequence is encoded by the essential region gene from plasmid pNBbl
XX used in the vector of the invention. The vector is a shuttle vector for a
XX Bifidobacterium, and contains: (a) a replicated essential region
XX originated from a plasmid pNBbl of Bifidobacterium breve ATCC 15698;
XX (b) a replicated essential region originated from a plasmid of E. coli;
XX and (c) an antibiotic-resistant gene originated from a plasmid of E. coli
XX and an antibiotic-resistant gene functioning by a Bifidobacterium. The
XX shuttle vector can be used to transform various Bifidobacteria species.
XX
SQ Sequence 494 AA;

Query Match 19.9%; Score 399.5; DB 19; Length 494;
Best Local Similarity 27.2%; Pred. No. 9,2e-31;
Matches 116; Conservative 64; Mismatches 169; Indels 77; Gaps 15;

4 VSAEHLGSG-KDRP-----PVLVSSPKR-----GIRHELRLPKLQOITTSFPN----- 44
38 VAAEAAASGARGSPPEKTKTKITPSSLRDRLRLAYGRRAESKILVRHAGSETLGFEPK 97
45 -----ACGRPLISGVNGVTITVNGPKSGGFGGLRSCGKWTCPCCAGKVGARADEISQVVA 99
98 LPRCARCGQPVDP--TGVGVMTNGEKARFTGTMLCGSIVMACPTCSAIIIRHRAHEVALAIG 155
100 HQL-----GTGSVAAVMTMTMRHTAQRHLDMTGLSA 131
156 NHAELRLRAAADQMAHEGQRLPELWVSDSFGVITFCTLTLRDRTWPLMTLDALIK 215
132 AAKATNGRWRTEREMWCGDGYRAVEITTHGKNGHVVHALLMPSGDSNILESFD 191
216 GWTKAINSGPMQRAEBRMKIRGFVRAIETTVNGHGHPIHVMPLDGLDDGQREBAMQ 275
192 AMFDEWTS--KLVSIGF-----AABLNRSGGLDVR-KIGAGDOVLAAYLTIKIASG 239
276 WLDLRWKTWVKAAYAKKKKQDGNPVYVAP-NDEHGIDLOFKSGKADGATAAAYITIKIQD 334
240 VG-----MEVSGDQKSGHGRNARAPETIANDAVG--DPQALIMREFPFGSGRRALAY 292
335 KGVVTLAOLRIARGLIKNGRMGVSVPQLDSSGLISDPQRBDMLIEWYQATLRRCITW 394
293 SRGLARAGLGAELTDAQIVQESAPVVA--IIPASMMMTITCQAPVYFEGILGLVEXG 351
335 SRGLKEDMEV--ELEDEDELAERKDELPGLVGVYVPRVYKDIRKSAPELTADALDAER- 452
QY 352 ATWENTL 357
```

```
DB 453 EDMQEV 458

RESULT 4
ID AAY97536 standard; Protein; 427 AA.
XX
XX AAY97536;
XX
DT 15-JAN-2001 (first entry)
XX
DE B. lactofermentum p48K protein sequence.
XX
KM Temperature sensitive plasmid; TSKCR; protein production;
XX temperature sensitive replication control region; p48K.
XX
OS Brevibacterium lactofermentum.
XX
PN EP1038966-A1.
XX
PD 27-SEP-2000.
XX
PF 16-MAR-2000; 2000EP-0105326.
XX
PR 16-MAR-1999; 99JP-0069896.
XX
PA (AJIN ) AJINOMOTO CO INC.
XX
PI Nakamura J, Kanno S, Kimura E, Matsui K, Nakamatsu T;
XX
DR WPI; 2000-573832/54.
XX
N-PSDB; AAA90934.
XX
PT Plasmids containing a temperature sensitive replication control regions
XX useful for breeding microorganisms for the production of amino acids by
XX fermentation -
XX
PS Claim 8; Page 14-17; 29pp; English.
XX
CC This sequence is the Brevibacterium lactofermentum p48K protein.
XX The invention relates to a plasmid containing a temperature sensitive
XX replication control region (TSRCR) and a marker gene (Mg). The TSKCR is
XX derived from plasmid pM4330 harboured by Brevibacterium lactofermentum
XX (ATCC 13869) and allows the plasmid to replicate autonomously at a low
XX temperature but does not allow the plasmid to replicate autonomously at
XX an elevated temperature in coryneform bacteria within a temperature
XX range in which the bacteria can grow. The plasmid can be used for
XX modifying a chromosomal gene in a coryneform bacterium, which may be used
XX for the production of useful substances, such as amino acids, by
XX fermentation, to change their genetic traits. Therefore, the plasmid can
XX be used for breeding microorganisms for the production of amino acids by
XX fermentation. The plasmid comprises a TSKCR that allows the plasmid to
XX replicate autonomously at an elevated temperature in coryneform
XX bacteria within a temperature range in which the bacteria can grow.
XX
SQ Sequence 427 AA;

Query Match 16.3%; Score 326; DB 21; Length 427;
Best Local Similarity 29.1%; Pred. No. 1.5e-23;
Matches 98; Conservative 58; Mismatches 131; Indels 50; Gaps 15;

34 LQOITTSFPNACGR-----PIGVNGVTITVNGPKSGGFGGLRSCGKWTCPCCAGV 86
27 MKTINSKALAGCHRRNRDEAVAVWS-----NG--ASQFGLNHSRMSPLAELEV 79
87 GAHRADEISQVVAHQLGTSVAVMTM-TMRHTAQRHLDMTGLSAAMKRAATNGRRWTE 145
80 MEERIEELAIATKHLAAGALIMFVGVTRHNRBSFPAQVEAGITAYASWVKTSQWKKE 139
146 REMYGGDGVAAVEITTHG-KNGMHVHVHALLMPSGDSVENILIESGSDAMFDRMYSKLYSL 204
140 RARYGVERTYSDYEVTDSWANGMHILRRMULFLDRPLSDELKAFEDSMFSRWSAGVKA 199
```


CC	acid (PHA) synthase, carotenoid biosynthesis enzymes, nitrile
CC	hydratases, ethylene forming enzyme, pyruvate decarboxylase, alcohol
CC	dehydratases, terpene synthases, and cholesterol oxidase in an
CC	Actinomycetales bacteria. The replication protein or plasmid stability
CC	protein are useful in cloning and expression vectors and particularly in
CC	shuttle vectors for the expression of homologous and heterologous genes
CC	in <i>Rhodococcus</i> sp. This sequence represents the <i>S. nigificans</i> pSM22
CC	replication protein described in the disclosure of the invention.
XX	
XX	Sequence 451 AA;
XX	
Query Match	16.1%; Score 322; DB 23; Length 451;
Best Local Similarity	30.8%; Pred. No. 4.1e-23;
Matches 101;	Conservative 41; Mismatches 110; Indels 76; Gaps 15
Db	
QY	49 PISGV-----NGVIVNGPKSGFGGIRSCGKMWICPCAGKYGARADEISQVAHQI 102
	3 PASGVITVAQRAAGISVLT-----GLMRCGRITWLCPCVCAATIRHRAEITAAVAVWI 54
Db	
QY	103 GTGSAV-VMTVMTHNTAGORLHDMTGISAAKAAATNGRR-----MRT----- 145
Db	55 KRGGTAIVLVPTFAHGHDTDLADMDALQGTREKADAPRRPGAYORLITGGTWAGRRAKD 114
QY	146 -----REMYGCGGYVAVAVEITGK-NGMVEVHVALMFGDV-----SENILESF 189
Db	115 GHRADDEBGRIDRIGYVGMIRATEVTVAGQINGMHPHIAVLVGGRTGERSAKQIVGTF 174
QY	190 --SDAMFDRMTSKLVSIGFAAPLR-----NSGIDVARKIGEAD-QVLAAYLTAKIA 237
Db	175 EPSBALDEWGGQRAV-WTALKKNAPQFTPDDRHGVDPRLETERANLAEVIATQ 233
QY	238 SG--VGMEVSGDGKSGRHGRAPWEI---AVDAVG--DPA-----LEIMREE 281
Db	234 DGAAPALEELARADIKTANGGVAPFELLGRIGDLTGWTEDDAGVSGLENNLARWHEYE 293
QY	282 FGSNGRAIAMSRLGRARAGIAGELTDA 309
Db	294 RATGGRRAIEMTRYLRQMLGLDGGDTGA 321
RESULT 8	
AA97537	
ID	AA97537 standard; Protein, 427 AA.
XX	
AC	AA97537;
XX	
DT	15-JAN-2001 (first entry)
XX	
DE	B. lactofermentum pSFK6 protein sequence.
XX	
KN	Temperature sensitive plasmid; TSCR; protein production;
KW	temperature sensitive replication control region; pSFK6.
XX	
OS	Brevibacterium lactofermentum.
XX	
FN	EP1038966-A1.
XX	
PD	27-SEP-2000.
XX	
PF	16-MAR-2000; 2000EP-0105326.
XX	
PR	16-MAR-1999; 99JP-0069896.
XX	
PA	(AJIN) AJINOMOTO CO INC.
XX	
PI	Nakamura J, Kanno S, Kimura E, Matsui K, Nakamatsu T;
XX	
DR	WPI; 2000-573832/54.
XX	
DR	N-PSDB; AAA90951.
XX	
PT	Plasmids containing a temperature sensitive replication control regions
PT	useful for breeding microorganisms for the production of amino acids by
PT	fermentation -

XX PS Disclosure; Page 19-22; 29pp; English.

XX CC This sequence is the Brevibacterium lactofermentum pSF6 protein.

CC CC The invention relates to a plasmid containing a temperature sensitive

CC CC replication control region (TSRCR) and a marker gene (MG). The TSRCR is

CC CC derived from plasmid pAM330 harboured by Brevibacterium lactofermentum

CC CC (ATCC 13869) and allows the plasmid to replicate autonomously at a low

CC CC temperature but does not allow the plasmid to replicate autonomously at

CC CC an elevated temperature in coryneform bacteria within a temperature

CC CC range in which the bacteria can grow. The plasmid can be used for

CC CC modifying a chromosomal gene in a coryneform bacterium, which may be used

CC CC for the production of useful substances, such as amino acids, by

CC CC fermentation, to change their genetic traits. Therefore, the plasmid can

CC CC be used for breeding microorganisms for the production of amino acids by

CC CC fermentation. The plasmid comprises a TSRCR that allows the plasmid to

CC CC replicate autonomously at an elevated temperature in coryneform

CC CC bacteria within a temperature range in which the bacteria can grow.

XX SQ Sequence 427 AA;

Query Match 15.9%; Score 318; DB 21; Length 427;

Best Local Similarity 28.8%; Pred. No. 9.6e-23;

Matches 97; Conservative 58; Mismatches 132; Indels 50; Gaps 15;

QY 34 LQOITTSSTFNACGR-----PISGVNGVTITVNGPKSGGFGILRSCGKGMICPCCAKRV 86

DB 27 MYKITSXALAGCHHRDEAVAVSWS-----NG--ASQFGLQNSHSRWSGLAELELV 79

QY 87 GAHRADETISQVVAHOLGTGVSVAWVTK--TMRHTAGORLHDLWTGLSAWKAATNGRRMTE 145

DB 80 MGERRIELAIATKTNHLAAGALMFMFVGVTRHNRSSQFAQVEGITYAVSSWTKSQMKKE 139

QY 146 REMYGCDDGVYRAVEITHG-KNGMHHVHALLMFSGDVSENILLESFSDAMFDRMTSKLYSL 204

DB 140 RARYGVHTYSYEVTDSMANGMHLHRNMLFLDRPLSDDELKAREDSWFSRWSAGVYKA 199

QY 205 GPAFLRNSG-GLD-VRKIGBADOVLAAYITKTASGVGMVSGSDGSGRRGNAPMEI 262

DB 200 GMDAPLRHGVKLDQVSTWGDPAK-MATYLA--GMSQELTGSATKTASRGSYTPFGM 255

QY 263 AVDAVGDPQALE-----LMREFEFGSMGRALAMRGIRARAGLGAELTDQI--- 311

DB 256 -LDMADQSDAGEMDAVLVARMRREYEVGSKNLR-SWSRG--AKRALGIDYIDADVRE 311

QY 312 VEOE-----ESAPVVAIIPARSMWMTIRT 335

DB 312 MEELYKLAGLEAPERVESTRAVALVPEDDMKLIQS 348

RESULT 9

AB06342

ID AB06342 standard; Protein; 427 AA.

XX AC ABB06342;

XX DT 29-MAY-2002 (first entry)

XX DE Brevibacterium lactofermentum p48K protein sequence SEQ ID NO:8.

XX KW Kanamycin resistant gene; mutation; high temperature resistance;

XX KM coryneform microbe; transformation.

XX OS Brevibacterium lactofermentum.

XX PN JP2002017362-A.

XX PD 22-JAN-2002.

XX PF 05-JUL-2000; 2000JP-0204236.

XX PR 05-JUL-2000; 2000JP-0204236.

XX XX

PA (AJIN) AJINOMOTO KK.

XX WPI; 2002-263243/31.

DR N-PSDB; ABL49734.

XX Transformation of high temperature-resistant coryneform microbe and

PT transformant -

XX Example; Page 22-23; 28pp; Japanese.

XX The present invention describes a method for transforming a high

CC temperature-resistant coryneform microbe in which the high temperature-

CC structure of cell wall of the microbe and an electric pulse is applied

CC to a solution containing the microbe having a changed structure of cell

CC wall and a DNA. Also described are: (1) transforming a high temperature-

CC resistant coryneform microbe in which the cell wall of a high

CC temperature-resistant coryneform microbe is weakened and electric pulse

CC is applied to a solution containing the microbe having weakened cell

CC wall and a DNA; (2) a transformant of a high temperature-resistant

CC coryneform microbe in which a recombinant DNA is introduced; and

CC (3) a transformant of a high temperature-resistant coryneform microbe

CC carrying both a plasmid vector derived from the high temperature-

CC resistant coryneform microbe and a plasmid vector derived from

CC coryneform glutamatum or Brevibacterium lactofermentum. The method

CC is used for transforming a high temperature-resistant coryneform microbe.

CC The present sequence represents a Brevibacterium lactofermentum p48K

CC protein sequence, which is used in an example from the present

CC invention.

XX SQ Sequence 427 AA;

Query Match 15.9%; Score 318; DB 23; Length 427;

Best Local Similarity 28.8%; Pred. No. 9.6e-23;

Matches 97; Conservative 58; Mismatches 132; Indels 50; Gaps 15;

QY 34 LQOITTSSTFNACGR-----PISGVNGVTITVNGPKSGGFGILRSCGKGMICPCCAKRV 86

DB 27 MYKITSXALAGCHHRDEAVAVSWS-----NG--ASQFGLQNSHSRWSGLAELELV 79

QY 87 GAHRADETISQVVAHOLGTGVSVAWVTK--TMRHTAGORLHDLWTGLSAWKAATNGRRMTE 145

DB 80 MGERRIELAIATKTNHLAAGALMFMFVGVTRHNRSSQFAQVEGITYAVSSWTKSQMKKE 139

QY 146 REMYGCDDGVYRAVEITHG-KNGMHHVHALLMFSGDVSENILLESFSDAMFDRMTSKLYSL 204

DB 140 RARYGVHTYSYEVTDSMANGMHLHRNMLFLDRPLSDDELKAREDSWFSRWSAGVYKA 199

QY 205 GPAFLRNSG-GLD-VRKIGBADOVLAAYITKTASGVGMVSGSDGSGRRGNAPMEI 262

DB 200 GMDAPLRHGVKLDQVSTWGDPAK-MATYLA--GMSQELTGSATKTASRGSYTPFGM 255

QY 263 AVDAVGDPQALE-----LMREFEFGSMGRALAMRGIRARAGLGAELTDQI--- 311

DB 256 -LDMADQSDAGEMDAVLVARMRREYEVGSKNLR-SWSRG--AKRALGIDYIDADVRE 311

QY 312 VEOE-----ESAPVVAIIPARSMWMTIRT 335

DB 312 MEELYKLAGLEAPERVESTRAVALVPEDDMKLIQS 348

RESULT 10

AAE13531

ID AAE13531 standard; Protein; 427 AA.

XX AC AAE13531;

XX DT 12-FEB-2002 (first entry)

XX DE B. lactofermentum p48K replication control region protein.

XX KW Coryneform bacteria; arginine repressor; L-arginine; liver function;

XX amino acid infusion; mutant; mutain.

```
XX OS Brevibacterium lactofermentum.
XX EN EPI154020-A2.
XX PD 14-NOV-2001.
XX PF 24-APR-2001; 2001EP-0109457.
XX PR 28-APR-2000; 2000JP-0129167.
XX PA (AJIN ) AJINOMOTO KK.
XX PI Suga M, Asakura Y, Mori Y, Ito H, Kurahashi O;
XX DR WPI; 2002-012660/02.
XX DR N-PSDB; AAD22583.
XX PT Coryneform bacterium with inactivated arginine repressor, useful for
XX PT fermentative production of arginine at high levels -
XX XX Example 1; Page 20-21; 30pp; English.
XX PS The present invention relates to Coryneform bacterium in which the
XX CC arginine repressor does not function normally and which produces
XX CC L-arginine. Coryneform bacterium is used to produce L-arginine which is
XX CC useful in agents for promoting liver function and in amino acid infusions
XX CC or in 'comprehensive' amino acid pharmaceutical. Coryneform bacterium
XX CC accumulate L-arginine to higher levels than wild-type strains, e.g. 120
XX CC compared with 20, mg/dl. The present sequence is
XX CC Brevibacterium lactofermentum p48K replication control region protein.
XX CC The replication control region DNA is used in the construction of shuttle
XX CC vector for Escherichia coli and Coryneform bacteria and temperature
XX CC sensitive vector.
XX XX
XX SQ Sequence 427 AA;
XX
XX Query Match 15.9%; Score 318; DB 23; Length 427;
XX Best Local Similarity 28.8%; Pred. No. 9.6e-23;
XX Matches 97; Conservative 58; Mismatches 132; Indels 50; Gaps 15;
QY 34 LQGITTSERFNACGR-----PISGVNCTIVNGKSGSGFGILSGCGKWCPCPCAGKY 86
DB 27 MKRTNISKALAGCHRRRDEAVAVSWSS-----NG--ASQFGLQNSHRSWSSLAELTV 79
QY 87 GAHRADEISQVVAHQLTGTSVAVMTM-TKRHTAGQRLHDLMTGLSAAMKAATNGRRWTE 145
DB 80 MGERIELAIAIKNHLAAGALMFMFGTVRHNRSQSFAYVEAGIKTAYSMTKTSQMKTE 139
QY 146 REMYGGDGYVRAVEITHG-KNGMHVVAHLLMFSGDVSENILSEPSDAMFDRWTSKLYSL 204
DB 140 RARYGVHEHTYSDYEVTDSWANGMHLRNMLLFLDRPLSDDELKAFEDSWFSRWSAGVVA 199
QY 205 GAAPLRNSG-GLD-VRKIGEADQVLAAYLKIASGVGMEVSGSGKSGRGNAPMWI 262
DB 200 GNDAPLRHEGVKLDQVTSWGDAK-MATYTLAK--GMSQELTGSATKPAASGYTFPMQ 255
QY 263 AYDAVAGDPOALE-----LMREPEFGSMGRRAIARSGLRARAGLAEITDAQI--- 311
DB 256 -IDMLADQSDAGEDMDAVLVARMRREVSGSKULRS-SWGRG--AKRALGIDYIDADVRE 311
QY 312 VEOE-----ESAPYVVAIIPARSWMMIRT 335
DB 312 MEEELYKLAGLEAPERVESTRAVALVKPDWMLIQS 348
XX
XX RESULT 11
XX ABB84281
XX ID ABB84281 standard; Protein; 456 AA.
XX AC ABB84281;
XX XX
XX DT 13-JAN-2003 (first entry)
```

```
XX DE S. lividans pJ101 replication protein.
XX XX Plasmid stability protein; replication protein; ethylene forming enzyme;
XX XX carotenoid biosynthesis enzyme; isoprenoid; pyruvate decarboxylase;
XX XX polyhydroxyalkanoic acid synthase; P4A synthase; nitrile hydratase;
XX XX alcohol dehydrogenase; terpene synthase; cholesterol oxidase;
XX XX shuttle vector.
XX OS Streptomyces lividans.
XX XX WO200255709-A2.
XX XX 18-JUL-2002.
XX XX 12-DEC-2001; 2001WO-US47868.
XX XX 12-DEC-2000; 2000US-254868P.
XX XX (DUPO ) DU PONT DE NEMOURS & CO E I.
XX XX Brannucci MG, Cheng Q, Kostichka KN, Tomb J;
XX XX WPI; 2002-557827/59.
XX XX
XX PT New nucleic acid molecule encoding replication protein/plasmid
XX PT stability protein, useful in cloning and expression vectors,
XX PT particularly shuttle vectors for expression of heterologous genes in
XX PT Rhodococcus species -
XX XX
XX PS Example 5; Fig 4A; 96pp; English.
XX CC This invention describes a novel nucleic acid encoding a replication
XX CC protein or a plasmid stability protein. The product of the invention is
XX CC useful for expression of nucleic acid such as genes encoding enzymes
XX CC involved in the production of isoprenoid molecules, polyhydroxyalkanoic
XX CC acid (PHA) synthases, carotenoid biosynthesis enzymes, nitrile
XX CC hydratases, ethylene forming enzyme, pyruvate decarboxylase, alcohol
XX CC dehydrogenase, terpene synthases, and cholesterol oxidase in an
XX CC actinomycetale bacteria. The replication protein or plasmid stability
XX CC protein are useful in cloning and expression vectors and particularly in
XX CC shuttle vectors for the expression of homologous and heterologous genes
XX CC in Rhodococcus sp. This sequence represents a Rhodococcus AN12 Rep
XX CC protein described in the disclosure of the invention.
XX XX
XX SQ Sequence 456 AA;
XX
XX Query Match 15.8%; Score 316; DB 23; Length 456;
XX Best Local Similarity 30.5%; Pred. No. 1.7e-22;
XX Matches 100; Conservative 40; Mismatches 112; Indels 76; Gaps 15;
QY 49 PISGV-----NGVTIVNGPKSGSGKSGKWCPCPCAGKGAHRADEISQVVAHQ 102
DB 3 PASGVIVAGTAAGTSVVL-----GLMRGRILVLCVPCATINHKRAEITTAAVENI 54
QY 103 GTGSVA-MVTMTRHTTAQQLHDLMTGLSAAMKAATNGRR-----WTE--- 145
DB 55 KKGGAAYIVTFFARHGHTDRDLADLDAQTEKTPDSRRPQAYORLITGTAWRARAD 114
QY 146 -----REMYGGDGYVRAVEITHGK-NGMHNVVAHLLMFSGDV-----SENILSEF 189
DB 115 GHRAADREGIRIRIVGVMTIRATEVTGQINGMHHAIVLVGRTGEGRSAGKIVATF 174
QY 190 --SDAMPDRWTSKLVSLGFAAPLR-----NSGGLDVRKIGED--QVLAAYLTGTA 237
DB 175 EPTGAALDEMGHMSV-WTALARKVNAFTPDDHGHVDPKRLTEFRDANDLAEYIAKQ 233
QY 238 SG--VGMEVSGDGSGSGHGNAPWEI---AYDAVGG---DEQA-----LEIMREEE 281
DB 234 DOKAPALIELARADLTATGAGGVAAPPPELLGRLIGDLTGKMTEDDAAGVSLMNLNRWHEY 293
QY 282 FSGMGRRAIARSGLRARAGLAEITDA 309
```

QY 289 ATAMSRGLKARAGJGAEIID----AQI VEQESAPVMVALLEAKRWMMIK 334

DB 117 SGNPKAFGAYQRMILAPAF IGRFEARRIKNDGIQIVKEAEDGIKRIIG IGMVKAABVI 1/8

QY 162 HG-KNGMHVHVALLMFSGDV-----SENILSFSDAMPDRYTSKL----- 201
 Db 177 RSKKNGYHPHLNLVFLGGELSTPAKGVGHFEPSETDLDDMEDWLMEMWAGALKRAD 236
 QY 202 ---VSLGPAAP---LRNSG-GLDYRKIGSGADQVLAAYLTG-----I 236
 Db 237 PKFEPSTDDDTFGCKCKGKGHGVMSIVASADVALILETLTKNQDKRERPSVDQDLGA 296
 QY 237 ASGVGMVSGSDGKSGR-HGNRAPEWIA-----VDAVSGDPQALE-----LMREF 280
 Db 297 AQAAMETARLDLSKTRGKSKMTFQILVRLMDIEVAGLDPDMAGGYGTPKQLRAMWQY 356
 QY 281 EFGSMGRRAIAMSRLRAPAGCA---ELTDQIVQEBSAPV 320
 Db 357 EEALAGRAIEMWTRGILRRHVDLDGDDDEETDLQYVEPEAPDL 399

RESULT 14

ID AAP93143 standard; protein; 314 AA.
 XX AAP93143;
 AC 25-MAR-2003 (updated)
 DT 03-OCT-2002 (updated)
 DT 29-JAN-1991 (first entry)
 DE Sequence encoded by ORF2 of plasmid pLH1 or pLAB1000.
 XX Alpha-amylase; lactic acid bacteria; carbohydrate.
 XX Enterococcus faecalis.
 OS Lactobacillus plantarum.
 XX WO8901970-A.
 PD 09-MAR-1989.
 PF 02-SEP-1988; 88WO-EP00813.
 XX 02-SEP-1987; 87EP-0401972.
 PR (PLAN-) PLANT GENETIC SYST.
 PA (UYLO-) UNIV CATHOL LOUVAIN.
 PA (UYLO-) UNIV CATHOLIQUE LOUVAIN.
 XX Michiels F, Delcour J, Mahillon J, Joos H, Platteuw C, Joosson K;
 PI WPI; 1989-085544/11.
 DR N-PSDB; AAN91248.
 XX
 PT Transformed lactic acid bacteria -
 PT using exogenous DNA high codes enzyme that breaks down
 PT oligosaccharide and/or polysaccharide, used in silage or as a
 PT probiotic
 PS Disclosure; Fig 3a; 112pp; English.
 XX
 CC The inventors claim an inoculum for silage which comprises lactic
 CC acid bacteria, esp. Lactobacillus plantarum and E. faecalis, transformed
 CC with an exogenous DNA encoding an enzyme which breaks down an oligosac-
 CC haride and/or a polysaccharide into a fermentable carbohydrate. The
 CC exogenous enzyme is especially an alpha-amylase encoded by the DNA
 CC sequence in AAN91246, particularly from nucleotides 184 to 2131 (Claims
 CC 2, 6 and 10). The transformed lactic acid bacteria can break down poly-
 CC saccharides and oligosaccharides in a silage crop to provide a complete
 CC source of the fermentable carbohydrates that the lactic acid bacteria
 CC need for producing lactic acid. Also claimed is a probiotic
 CC characterised by the transformed lactic acid bacteria which can be used
 CC for establish-ing and maintaining optimal intestinal flora in animals,
 CC including humans. A plasmid contg. the DNA sequence in AAN91246 is
 CC claimed. The plasmid is pLH1 or pLAB1000.

CC (Updated on 03-OCT-2002 to add missing OS field.)
 CC (Updated on 25-MAR-2003 to correct PA field.)
 SQ Sequence 314 AA;

Query Match 6.7%; Score 134; DB 10; Length 314;
 Best Local Similarity 22.4%; Pred. NO. 0.00012;
 Matches 56; Conservative 50; Mismatches 106; Indels 38; Gaps 10;

QY 75 KGWICPCGAGVGAHRADEISQV--AH-QLGTGSVAAMVTMTHTAGQRLHDLWTGISA 131
 Db 69 KSRCLPCLNMRMSMQSQMLQVDEAHKQKRTGFLFTITABASSENILKQEVKRGK 128
 QY 132 AKKATNRRMRTEREMTCGQYPAVEITGKNG-MHVVHVALLMFSGDVSENILSFPS 190
 Db 129 A-----ISKLPQYKPAKNLIGYRSTETITKNGYTHQHWHVLFVPTFKDSANYIN 183
 QY 191 DAMPDRMTSKVSLGAPALPNSGGLDYRKIGSGADQVLAAYLTITAGVGMVSGDGCK 250
 Db 184 DAEMSKLMXKRAKLDY-XPIYVBAVRSNKAKGNKSLTASQET----- 226
 QY 251 SGRHGNRAPWEI-AVDVAGDPOA-LELMREFEFGSMGRRAIAMSRLG--PARAGLGAEL 306
 Db 227 -----AKYQYKSDILITNDQERDLQYVEDLEQGLASRQISYQ-GLFKETRKQLQLED 278
 QY 307 TDAQIVQEBS 316
 Db 279 VDAMLINVD 288

RESULT 15

ID ABUS6572 standard; Protein; 647 AA.
 XX ABUS6572;
 AC 02-APR-2003 (first entry)
 DT Lung cancer-associated polypeptide #165.
 DE Lung cancer-associated polypeptide #165.
 XX
 KW Lung cancer-associated polypeptide; cytostatic; emphysema;
 KW antiinflammatory; antiasthmatic; non-small cell lung cancer; atelectasis;
 KW small cell lung cancer; benign lesion; precancerous lesion; bronchitis;
 KW chronic obstructive pulmonary disease; hypersensitivity pneumonitis;
 KW interstitial pulmonary fibrosis; fibrosis; asthma; bronchiectasis.
 XX Undenitified.
 OS
 XX WO200286443-A2.
 FN 31-OCT-2002.
 PD 18-APR-2002; 2002WO-US12476.
 PF 18-APR-2001; 2001US-284770P.
 PR 10-MAY-2001; 2001US-290492P.
 PR 09-NOV-2001; 2001US-339245P.
 PR 13-NOV-2001; 2001US-350666P.
 PR 29-NOV-2001; 2001US-334370P.
 PR 12-APR-2002; 2002US-372246P.
 XX (EOSB-) EOS BIOTECHNOLOGY INC.
 PA Aziz N, Murray R;
 PI WPI; 2003-093161/08.
 DR N-PSDB; ABX76501.
 XX
 PT Detecting a lung cancer-associated transcript in a cell from a patient
 PT for treating lung cancer, by contacting a biological sample from the
 PT patient with a polynucleotide that exhibits increased or decreased
 PT expression in lung cancer -

PS Claim 27; Page 315; 453pp; English.

CC The invention relates to a method for detecting a lung cancer-associated
CC transcript in a cell from a patient, comprising contacting a biological
CC sample from the patient with a polynucleotide that selectively hybridises
CC to a sequence that is at least 80 % identical to a gene that exhibits
CC increased or decreased expression in lung cancer samples. Lung
CC cancer-associated polynucleotides and polypeptides are used for
CC identifying a compound that modulates a lung cancer-associated
CC polypeptide, for inhibiting proliferation of a lung cancer-associated
CC cell to treat lung cancer in a patient and for treating a mammal having
CC lung cancer by administering a modulatory compound identified. The
CC methods are useful for treating lung cancer, such as small cell lung
CC cancer, non-small cell lung cancer or other benign or precancerous
CC lesions, e.g. atelectasis, emphysema, bronchitis, chronic obstructive
CC pulmonary disease, fibrosis, hypersensitivity pneumonitis, interstitial
CC pulmonary fibrosis, asthma and bronchiectasis. The genes, polynucleotides
CC and polypeptides are useful for diagnostic purposes and as targets for
CC screening for therapeutic compounds that modulate lung cancer, such as
CC antibodies. Sequences AB056408-AB056745 represent lung cancer-associated
CC polypeptides of the invention.

Sequence 647 AA;

Query Match	5.1%;	Score 103;	DB 24;	Length 647;
-------------	-------	------------	--------	-------------

Best Local Similarity 20.8%; Pred. No. 0.41;

Matches 86; Conservative 37; Mismatches 150; Indels 140; Gaps 18;

QY	2	TSVSAHLSHGKRPRLVSSDKRGRIHRLRPLDQITTSFTFNAAG-----RPLSGV	55
Db	120	TGSGRAVBSGDAFDTA-----LRRPAGIQVSS-SACGASSENDAVRRVPL	168
QY	54	NGVTIVNGRGS-----GFGGL-----RSCG----	74
Db	169	AGALARAGRRRTTHCRPCWLLGLGSLDPAPIYHEAAGRGGLHPRWGAQHRCGRRA	228
QY	75	-----KGNICPCCAKGVANHRADEISOVVAHQLTGSSVA---MVTWMTHTA	118
Db	229	RCARAPAGRPBARGLQRPAPVLGRGTAQ-----APLHPGERAPAGILAVLPRR	279
QY	119	GGALHDLMTG-----LSAAMKATNTGRR---RTREMYGCDGYVRAVEITHGNKNHTH	170
Db	280	SRRRHAIVGGGAFETLLHRAEMRGTPQHRGRARSWKEN-----RCH	320
QY	171	WHAALMFSGVSENILSFSDAMPFRWTSKLVSLFAFAPLR-NSGGLDVRKIGGEADVTL	229
Db	321	LRA-----NGYLCKYQYEVLCRAPRPGAAS---NLSTYRAPFQJHSAALDFSPGIEVALC	373
QY	230	AAVYLTKIASGVGMEVG-----SGD-----GKSGRHNRAPWELIANDVAGDPEQALEW	277
Db	374	RQGLPISVTCIADEIGARDKLSGDVLCPCRPYLRACKCAELPNCIDLDIG--FACECA	431
QY	278	REPEFSGMGRALAMSRGLPARAGIGAEILTDAQIYEQESAPVVAITPARSW	330
Db	432	TGELIGDGDGSCVTSGEQPTLGGTG-----VPTRRPATATISVPQGTW	476

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 8, 2003, 23:04:13 ; Search time 27 Seconds
(without alignments)
593.919 Million cell updates/sec

Title: US-10-007-527A-2
Perfect score: 2005
Sequence: 1 MTSVSAEHLGKDRPPPLVS.....HLHYRLPADVRPPIISVRK 379

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-Processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA:*
1: /cgn2_6/prodata/2/iaa/5A COMB.pep:*
2: /cgn2_6/prodata/2/iaa/5B COMB.pep:*
3: /cgn2_6/prodata/2/iaa/5A COMB.pep:*
4: /cgn2_6/prodata/2/iaa/5B COMB.pep:*
5: /cgn2_6/prodata/2/iaa/5A COMB.pep:*
6: /cgn2_6/prodata/2/iaa/5B COMB.pep:*

Pred. No. is the number of results predicted by chance a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	326	16.3	427	US-09-521-668B-18	Sequence 18, Appl
2	318	15.9	427	US-09-521-668B-20	Sequence 20, Appl
3	101.5	5.1	882	US-09-252-991A-17653	Sequence 17653, A
4	101	5.0	1395	US-09-252-991A-30345	Sequence 30345, A
5	99.5	5.0	676	US-09-252-991A-23181	Sequence 23181, A
6	98.5	4.9	621	US-09-252-991A-19231	Sequence 19231, A
7	98.5	4.9	1409	US-09-252-991A-22282	Sequence 22282, A
8	97.5	4.9	1245	US-09-252-991A-30935	Sequence 30935, A
9	97	4.8	713	US-09-315-409-11	Sequence 11, Appl
10	97	4.8	713	US-09-568-102-11	Sequence 11, Appl
11	97	4.8	713	US-09-567-969-11	Sequence 11, Appl
12	97	4.8	713	US-09-568-480-11	Sequence 11, Appl
13	97	4.8	713	US-09-568-486-11	Sequence 11, Appl
14	97	4.8	713	US-09-568-472-11	Sequence 11, Appl
15	97	4.8	713	US-09-567-899-11	Sequence 11, Appl
16	97	4.8	1403	US-09-252-991A-25500	Sequence 25500, A
17	95.5	4.8	626	US-09-252-991A-30863	Sequence 30863, A
18	95.5	4.8	2890	US-09-413-814-67	Sequence 67, Appl
19	95.5	4.8	3798	US-09-335-409-6	Sequence 6, Appl
20	95.5	4.8	3798	US-09-568-102-6	Sequence 6, Appl
21	95.5	4.8	3798	US-09-567-969-6	Sequence 6, Appl
22	95.5	4.8	3798	US-09-568-480-6	Sequence 6, Appl
23	95.5	4.8	3798	US-09-568-486-6	Sequence 6, Appl
24	95.5	4.8	3798	US-09-568-472-6	Sequence 6, Appl
25	95.5	4.8	3798	US-09-567-899-6	Sequence 6, Appl
26	94.5	4.7	429	US-09-252-991A-31382	Sequence 31382, A
27	94	4.7	414	US-09-252-991A-22296	Sequence 22296, A

28	92.5	4.6	579	US-09-198-452A-918	Sequence 918, App
29	92.5	4.6	612	US-09-252-991A-29283	Sequence 29283, A
30	92	4.6	671	US-09-252-991A-31862	Sequence 31862, A
31	91	4.5	158	US-09-252-991A-16612	Sequence 16612, A
32	91	4.5	561	US-09-252-991A-16726	Sequence 16726, A
33	91	4.5	1174	US-09-252-991A-29279	Sequence 29279, A
34	90.5	4.5	611	US-09-252-991A-25119	Sequence 25119, A
35	89.5	4.5	415	US-09-252-991A-25764	Sequence 25764, A
36	89.5	4.5	601	US-09-252-991A-30589	Sequence 30589, A
37	89	4.4	503	US-09-252-991A-23522	Sequence 23522, A
38	89	4.4	881	US-09-252-991A-31702	Sequence 31702, A
39	89	4.4	3562	US-09-679-279-14	Sequence 14, Appl
40	89	4.4	3739	US-09-320-878-2	Sequence 2, Appl
41	89	4.4	3739	US-09-105-537-33	Sequence 33, Appl
42	89	4.4	3739	US-09-141-908-3	Sequence 3, Appl
43	89	4.4	3739	US-09-657-440-2	Sequence 2, Appl
44	89	4.4	11877	US-09-105-537-6	Sequence 6, Appl
45	88.5	4.4	514	US-09-252-991A-30952	Sequence 30952, A

ALIGNMENTS

```
RESULT 1
US-09-521-668B-18
; Sequence 18, Application US/09521668B
; Patent No. 6303383
; GENERAL INFORMATION:
; APPLICANT: NAKAMURA, JUN
; APPLICANT: KANNO, SOHEI
; APPLICANT: KIMURA, EICHIRO
; APPLICANT: MATSUI, KAZUHIKO
; APPLICANT: NAKAMATSU, TOSYOSHI
; TITLE OF INVENTION: TEMPERATURE SENSITIVE PLASMID FOR CORYNEFORM BACTERIA
; FILE REFERENCE: 0010-1093-0
; CURRENT APPLICATION NUMBER: US/09/521,668B
; CURRENT FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: JP 11-69896
; PRIOR FILING DATE: 1999-03-16
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 18
; LENGTH: 427
; TYPE: PRT
; ORGANISM: Brevibacterium lactofermentum
US-09-521-668B-18

Query Match      16.3%; Score 326; DB 4; Length 427;
Best Local Similarity 29.1%; Pred. No. 3.7e-25;
Matches 98; Conservative 58; Mismatches 131; Indels 50; Gaps 15;

QY 34 LQGITSEFNNACGR-----PISGVNGVTIVNGPGSGPGGLRSCGKGMICPCGAKV 86
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 27 MYKIMSKALACHKRKRREDAVAVSS-----NG--AQFGLQSHSRMSPLAELEV 79
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 87 GAHRADEISQVVAHQIGTSSVAVVTN--TNRHTAGQRLHDLMTGLSAMKRAATNGRRKTE 145
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 80 MGRRIETLAIATKRNHLAAGALMMFVGTVRHRSQFAQVEAGIKTAYSSMVTQSKKE 139
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 146 RMYGCDGVYRAVEITHG-KNGMNVVHALLMPSGVSEKILESBDANFDRSTSLVSL 204
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 140 RARYGEHYHSYDEVDSWANGMHLRNNLLFLDRPLSDDELAFDSMFSRWSAGVYKA 199
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 205 GFAPAPRNAG-GLD-VRKIGEDAVLYALVLTIASGVGEVSGSGKSGRHNRAPEI 262
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 200 GMDAPAREHGVLDQVSTYMGDAK--MAYIAK---GMOQLTGSATKTRASKSYTFPQM 255
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 263 AYDAVGDDPOALE-----LMREFPSMGRRAIAMSGLRARGAGELTDAOI--- 311
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 256 -LDMTLADQSDAGMDAVLVARWRREYVSKNIRS--SWSRG--AKRALGIDYIDAVRRE 311
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 312 VBOE-----ESAPVYVAIIPRSMWMTIT 335
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

Db 312 MEEELYKLAGLEAPERVESTRAVALVPEDDMKLIQS 348

RESULT 2

US-09-521-668B-20
Sequence 20, Application US/09521668B

Patent No. 6303383

GENERAL INFORMATION:

APPLICANT: NAKAMURA, JUN

APPLICANT: KANNO, SOHEI

APPLICANT: KIMURA, EICHIRO

APPLICANT: MATSUI, KAHIKO

APPLICANT: NAKAMATSU, TOSYOSHI

TITLE OF INVENTION: TEMPERATURE SENSITIVE PLASMID FOR CORYNEFORM BACTERIA

FILE REFERENCE: 0010-1093-0

CURRENT APPLICATION NUMBER: US/09/521,668B

PRIOR FILING DATE: 2000-03-08

PRIOR FILING DATE: 1999-03-16

NUMBER OF SEQ ID NOS: 20

SOFTWARE: PatentIn version 3.0

SEQ ID NO 20

LENGTH: 427

TYPE: PRT

ORGANISM: Brevibacterium lactofermentum

US-09-521-668B-20

Query Match 15.9%; Score 318; DB 4; Length 427;
Best Local Similarity 28.8%; Pred. No. 2,5e-24;
Matches 97; Conservative 58; Mismatches 132; Indels 50; Gaps 15;

34 LQOITTFETFNACGR-----FISGVNGVITVNGPKSGFGGLRSCGKMTCPCCAGKY 86

27 MKTITSKALAGCHRRRDEAVAVSWS-----NG--ASQFEGELQSHSRMSSSLAELEV 79

87 GAHRADISQVVAHQGTSSVAMVTM--TKRHTAGQHLHLMTGLSAAWKAATNGRRKRE 145

80 MGRRIEILAIATGNHIAAGGALMMPVGTVAHNRSCQFAQVEAGIKTAYSVMYTSQWKKE 139

146 REMYGGDGYVRAVEITHG--KNGWHVHVHALLMPSGVSNTLESFSDAMFDRWTSGLVSL 204

140 RARYGVHTYSDEYVDSNANGWHLRNMILLFDRLPLSDDELAFEDSMFSRWAGVVA 199

205 GFAPPLRNSG-GID-YRKIGSEADQVLAAYLTIKIASGVMEVSGSGKGRHGNRAPEWI 262

200 GMDAPLREHGVKLDQVSTWGDAAK--MATYIAK---GMSGELGSAITKASKSYPPFQM 255

263 ANDAVGDDPOLAF-----LMREFPFGSGMERRALANSRGLRPARGLGAELTDAOI--- 311

256 -IDMTLADQSDAGEDMDAVLVARRREYEVSKNLR--SWSRG--AKRALGIDYTDADVRR 311

312 VEOE-----ESAPVVAIIIPASMMWIRT 335

312 MEEELYKLAGLEAPERVESTRAVALVPEDDMKLIQS 348

Db 312 MEEELYKLAGLEAPERVESTRAVALVPEDDMKLIQS 348

Db 312 MEEELYKLAGLEAPERVESTRAVALVPEDDMKLIQS 348

Db 312 MEEELYKLAGLEAPERVESTRAVALVPEDDMKLIQS 348

Db 312 MEEELYKLAGLEAPERVESTRAVALVPEDDMKLIQS 348

Db 312 MEEELYKLAGLEAPERVESTRAVALVPEDDMKLIQS 348

Db 312 MEEELYKLAGLEAPERVESTRAVALVPEDDMKLIQS 348

Db 312 MEEELYKLAGLEAPERVESTRAVALVPEDDMKLIQS 348

Db 312 MEEELYKLAGLEAPERVESTRAVALVPEDDMKLIQS 348

Db 312 MEEELYKLAGLEAPERVESTRAVALVPEDDMKLIQS 348

LENGTH: 882

TYPE: PRT

ORGANISM: Pseudomonas aeruginosa

FEATURE:

NAME/KEY: UNSURE

LOCATION: (730)

OTHER INFORMATION: Identity of amino acid at the above locations are unknown.

US-09-252-991A-17653

Query Match 5.1%; Score 101.5; DB 4; Length 882;
Best Local Similarity 21.7%; Pred. No. 0.17;
Matches 87; Conservative 48; Mismatches 125; Indels 141; Gaps 20;

47 GRPISGVNGVTTVNGPKSGFGGLRSCGKMTCPCCAGKY 98

487 GRVLGVGLALBOGALGNR-----VAFVVAQGRKVGIEAERQSRPWHRRPTE 535

99 -----AQIGTGS-----VAMVTMTHTAGORLHDL-----W 126

536 IRHAPADGVGGAGRAVELGGGQQLLQRLVVQVRIAL--HDQQRAGVGGRGHRGTGL 594

127 TGLSAAMKATNGRRWTEREMYGCDGYRA-----VEITH-----GKNGMHV 169

595 VGVAAGHGAVDQAAGRGDAPVLDPAAVVALAVLVASGHCQPYAFQVRLEVGCGGHA 654

170 HV-----HALMFGDVSNTLESFSDAMPD-----RMTSKLVS--LGFAPPL 210

655 GVGVAAVAGAEVDVHAL---AGDAGGVQAPAGAPVVLIGGVQAVEGLVADVLGLAAPT 711

211 -----NSGGLDVPRKIGSEADQVLAAYLTIKIASGVMEVSGSGKGRHGNRAPEWI 260

712 VDRPHGVGQRLVGGLEIKRVRRAEQAVAL-----VAGDVHLCKVGH----- 756

261 EIAVDVAVG-GDPOLAEIWM-----REFPFGSGMERRALANSRGLR-ARAG 301

757 PVHADAVARGAAGDADVAVGVIVAHRAADARAANDVIGAAGRRLVABGSGIMARVE 816

302 LGAEILTDAQIYQESAPVMAIIPASMMWIRTCAPIYFG 342

817 TGVHLADLHALADAGIGLVGHPHFOAPVILL-----VFG 851

Db 817 TGVHLADLHALADAGIGLVGHPHFOAPVILL-----VFG 851

Db 817 TGVHLADLHALADAGIGLVGHPHFOAPVILL-----VFG 851

Db 817 TGVHLADLHALADAGIGLVGHPHFOAPVILL-----VFG 851

Db 817 TGVHLADLHALADAGIGLVGHPHFOAPVILL-----VFG 851

Db 817 TGVHLADLHALADAGIGLVGHPHFOAPVILL-----VFG 851

Db 817 TGVHLADLHALADAGIGLVGHPHFOAPVILL-----VFG 851

Db 817 TGVHLADLHALADAGIGLVGHPHFOAPVILL-----VFG 851

Db 817 TGVHLADLHALADAGIGLVGHPHFOAPVILL-----VFG 851

Db 817 TGVHLADLHALADAGIGLVGHPHFOAPVILL-----VFG 851

Db 817 TGVHLADLHALADAGIGLVGHPHFOAPVILL-----VFG 851

Db 817 TGVHLADLHALADAGIGLVGHPHFOAPVILL-----VFG 851

Db 817 TGVHLADLHALADAGIGLVGHPHFOAPVILL-----VFG 851

Db 817 TGVHLADLHALADAGIGLVGHPHFOAPVILL-----VFG 851

Db 817 TGVHLADLHALADAGIGLVGHPHFOAPVILL-----VFG 851

Db 817 TGVHLADLHALADAGIGLVGHPHFOAPVILL-----VFG 851

Db 817 TGVHLADLHALADAGIGLVGHPHFOAPVILL-----VFG 851

```

Db      569 HOSIQRKRG- EAVVAAA-VGGDLADRVAAVIGGHAARLGAALRGAVAGVD- -H 622
QY      117 TAGQ--RLHDLWTGLSAAMKAAATNG-----RRMTEREMTGC-----DGYV 155
Db      623 WRGGRVAVH---GDTAARTGAVAGVGVDHAGVGRGSEAPVAALIGDLPDGA 678
QY      156 RAVEITTEGKNGHVVHVALIMFSGDVSENILESFDAMFDRMTSKLVSLGFAPLR- -211
Db      679 VAIGQGHG-----GARLGAALRGAVARVDHNRGRCHGVHGDAGRTGVA 723
QY      212 ---NGGLDVYKIG-----GEADQVLAALVTLTASGVGMEVSSGGKSGRHNRAPMELA 263
Db      724 GRVGRGVDRHRAVGGRRARGEG-PVAAAAGDLANRVAAVQQGHG-SARLGAALRGRTV 781
QY      264 VDAVAGDDPOLALIMREFEFGSMGRR-ALAW-----SRGLR-----ARAGLGAE 305
Db      782 ARIDGRGR-----RDGIGNAGKRTAVAGGIPGHVHDGHTVROGGRERERVAVIGGG 836
QY      306 LTDAQIVGQESAPYVVALIPARSMWMTITCAPYVFGELIGVEAGATWENLRDHLHYRL 365
Db      837 LAD-----RIMAVGQGHAAARFGSPAPPAVARL-----DHRVYR 873
QY      366 PAAD 369
Db      874 HAVD 877

```

```

RESULT 5
US-09-252-991A-23181
; Sequence 23181, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 23181
; LENGTH: 676
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-23181

```

```

Query Match      5.0%; Score 99.5; DB 4; Length 676;
Best Local Similarity 25.6%; Pred. No. 0.18;
Matches 89; Conservative 31; Mismatches 135; Indels 93; Gaps 19;

```

```

QY      62 PRGSGFGILRSCGKMTIC--PCCAGKVAHRADEIS-----QVAHQLTGGSVAMVTMT 113
Db      318 PRGPAAGRRASRGCGCRGCGAGRTAGRASPVATGRSGCCSRAPGAPVA----- 372
QY      114 MHHTAGQRLHDLMTGLSAAMKAAATNRERKTEREMTGCIGYRAVEITTEGKNGHVVHVA 173
Db      373 -GRGAENRRRRSRWSLPAS--CLSRGR-----GVCSVARRRPGG----- 409
QY      174 LMFSGDVSENILESFDAMFDRMTSKLVSLGFAPDLRNSGGLDVKIGSEADQVLAALV 233
Db      410 -AVRGSVSCALLTQAKRAGD-----AVDLGIGA-----YLLDQRFBENARQFLAAQR 458
QY      234 TTIASGVG-MEVGS-----GDGK-----SGRHGNRAPEIADV-AVGGD-PQALBLME 279
Db      459 GRRRAAFETEGALPFOAHGEGRLDVRQAGPAQFGFAYVDHQAAGDVAAALE----- 514
QY      280 PFGSGNR--RALMSRGLRARAAGLTAQIVGQESAPV--MVAIIPARSMWMTIR 334
Db      515 -----GRWVQATEQTEGVAE-----GQDVGEAQVAAQRRGPAHRSAAVVVDDAQSHHLLA 564

```

```

QY      335 TCAPYV-----FGELIGVEAGA--TWENTLDHLHYRLPAAD 369
Db      565 GRAALVEHLEDHVPPRHDEGQQQAAGAGELAEERQAEHQHQDQAD 612

```

RESULT 6

```

US-09-252-991A-19231
; Sequence 19231, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 19231
; LENGTH: 621
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-19231

```

```

Query Match      4.9%; Score 98.5; DB 4; Length 621;
Best Local Similarity 22.4%; Pred. No. 0.21;
Matches 62; Conservative 27; Mismatches 87; Indels 101; Gaps 11;

```

```

QY      175 LMFSGDVSENILESFDAMFDRW-----TSKLVSLGFAPLRNSGL 216
Db      323 LTVSGVDVSNLIGSLTDQLQRFHCLRGDESLLPHRLQFELRQEFMLRRAGV- 381
QY      217 DVKRTGSEADQVLAALVTLTASGVGME-----VSGDGKSGRHGNRAPEIADVAGDQ 272
Db      382 ---NFGATRMSLLMYQRLADGIGEGYHCYTLGEG-----ADELPWGYR 425
QY      273 ALEMRPEFGMGRRATAMSGRLARAGLGAELTDAQIVGQ-ESSA----- 318
Db      426 HLEMRRDAPPRFPAAWGEYRKALAEPRGRVAREIEELAEHALQGLEAIG 485
QY      319 --PVWVAIIP--ARSMWMTITC--APYVG-----ETLG-----LVE 349
Db      486 QFDLHYSLPELRRADHLLMSRTIARPTYLHGALAQARGLQRIVDPAKAPVALLAQ 545
QY      350 AGATWE-----NLRDHLHYRLPA 367
Db      546 REKRWQAQPKRHFRLLPFRERWPQALGEMRHLAERLPA 582

```

RESULT 7

```

US-09-252-991A-22282
; Sequence 22282, Application US/09252991A
; Patent No. 6551795

```

GENERAL INFORMATION:

```

; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 22282
; LENGTH: 1409
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa

```

US-09-252-991A-22282

Query Match 4.9%; Score 98.5; DB 4; Length 1409;
Best Local Similarity 23.8%; Pred. No. 0.69;
Matches 103; Conservative 51; Mismatches 160; Indels 119; Gaps 24;

QY 14 RPPVLSVSRGRIRHELR--PKIQITSETFNACGRPIISGVNGVTI-----VNGP 62
DB 275 RPPALDQASGGILQVQRPCFGLQLSR---RLASRDLAGPAGETPRDPHGLILCY 330
QY 63 KSGS-----FCGRSGCKMIC--PCCAG---KVGARADEISQVVAHQ 101
DB 331 RQSGATDHPVAGHALLFEITFGQIQAIPEVVAADPHLALQKQKAGARAEITTCACAQIQ 390
QY 102 L-----GT---GSVAMVTMTMRHTAGCR--LHDLWTGLSAWKATN----- 138
DB 391 LIAIAGLRQPTQGVQGVAAATAQSLAVGQQRRCGRVEGTALISQAVAAAGNPEAVAV 450
QY 139 GRP---WRTEREMYGCDGYVRAVEITHGKNGMHVHALLMPSGDVSENILSFSQAMD 195
DB 451 GQRIIVLRQQQVAPQAGIOPVAVAHQDA-----VFAGIDETI-----DRPVR 496
QY 196 RMTSKVSLGFAPLNSGLDVRKIGBADQVLAAYLTIKIASGVMEVSGDGKSGRHG 255
DB 497 RLRQHVAGLQAAAMGQDVGVD--RHVVGQD--CLAI-----GQQRADVADALAGRH 548
QY 256 NEAPEIADVAVGDPQ--ALELMREFEGSMGRRAIAMSRLRARAGLGAELTDAQIV 312
DB 549 QAAQLE-AVEQFVAVQAALALGHHVSGRLGRLAGTDHDLLEAAGIGAE---SVL 603
QY 313 EDEESAPVWVAIIP-----ASMMIRTCAPVFEIILGLVBAQTWE-----NLR 358
DB 604 QQAALAHVVGVAHPDQLCAARRGIL-----RLVVAVQAG-LMRFGLPLAAGVA 652
QY 359 DHL--HYRLPAAD 369
DB 653 DHLAVEPRIGAAD 665

RESULT 8

US-09-252-991A-30935
; Sequence 30935; Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 30935
; LENGTH: 1245
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-30935

Query Match 4.9%; Score 97.5; DB 4; Length 1245;
Best Local Similarity 21.2%; Pred. No. 0.73;
Matches 86; Conservative 50; Mismatches 143; Indels 127; Gaps 19;

QY 9 LSGKDP-PYLVSDKRGIRHELRPKIQITSETFNACGRPI-----SGVNGVT 57
DB 188 LRQPRSGPAPFGRQRPGRHRTYPVQ-----RPAQAPRRGGGTGARVA 235
QY 58 IVNGPKSGFGGKRSCKGM-----ICPCAGKVGARADEISQVVAHQ 102
DB 236 AVEGE-----GCRITCAAGKQRAFPGRRAAPGTRPCGAGAGRQLAMATGVRQAAGRT 269

QY 103 GTGSVAMVTMT-----MRHTAGRLHDL-----WTGLSAWKATNCR 140
DB 290 GGGVIALSRRADPPGRSPACQAGIRYGADRPFRFSPRPAALAGAAAGW----- 342
QY 141 RMRTEREMYGCD-GYVRAVEITHGKNGMHVHALL-----MPSGDVSENILSFS 190
DB 343 -WR-----GADPGALRRV-VSINENIFSHMMVFLPAASSLSVAAGDIQQSIGESS 394
QY 191 DAM-FDR--WTSKVLISGFAPLNSGGLDVRKIGBADQVLAAYLTIKIASGVMEVSG 247
DB 395 MTLPTTAAAMPPLCSAAVLGALMWAASAAERRFDLPQPLAASLSRLAQOAVQVLF 454
QY 248 DKSGRHGNRAPWEIADVAVGDPQALE---LMREFEGSMGRRAIAMSRLRARAGLGA 304
DB 455 R--SLRLGRAP--ALSQSYGVREALERLLVGSSELEVEADGYVARRQVDAYSQNAL 509
QY 305 ELTDQIV-----EDESAPVWVAIIP 326
DB 510 QL-DAQITVGNRGREVDASNGRSTLTRDIERQADNIPSLQITIP 554

RESULT 9

US-09-335-409-11
; Sequence 11; Application US/09335409
; Patent No. 6121029
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goettlich, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/335,409
; CURRENT FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 11
; LENGTH: 713
; TYPE: PRT
; ORGANISM: Sorangium cellulosum
US-09-335-409-11

Query Match 4.8%; Score 97; DB 3; Length 713;
Best Local Similarity 22.1%; Pred. No. 0.36;
Matches 78; Conservative 38; Mismatches 157; Indels 80; Gaps 12;

QY 30 LRPKIQITSETFNACGRPIISGVNGVTIYNGPKSGFGGKRSCKGMICPCAGRYGAH 89
DB 302 MRVDVSLQRTTPAPAMGTVALLLATATPAKVPALGRLGLKSGSEALVALVGLMKG-- 359
QY 90 RADEISQVVAHQIGTGS-----VAMVTMTMRHTAGRLHDLWTGLSAWKATNCR 140
DB 360 -TDLIAIVGVELGILSNAYTMVAVVAVLV---TABPALL-ITLKR--PPTQER 410
QY 141 RMRTEREMYGCDGYVRAVEITHGKNGMHVHALLMPSGDVSENILSFSQAMDERTSK 200
DB 411 SARLREEARARAYIPGVE---RIVPIVAHALPEFANDIVESIVAS----- 454
QY 201 LVSLGFAPLNSGGLDVRKIGBADQVLAAYLTIKIASGVMEVSGDGKSGRHNR--- 257
DB 455 -----KRLGRTVD-ITELSVQQAQPGSPRAAGASGLAALGARLRV 496
QY 258 APEIADVAVGDPQALELMREFEGSMGRRAIAMSRLRARAGLGAELTDAQIVQEES 317
DB 497 GTRQKRELRGSIQALIRSRDLDLVIGARSPARAKGSF-----GRLODAVQRAESN 551
QY 318 APVVAALIPASMMMTITQCAPVFGIILGLVEAGATMENIRDLHTRLPADV 370
DB 552 VLVVGDPPAAE---RASRRITLVPIG-----LEYSFAAADL 586

```

RESULT 11
US-09-567-969-11
; Sequence 11, Application US/09567969
; Patent No. 6355457
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goeschlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/567,969
; CURRENT FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0

```

	Query Match	4.8%;	Score 97;	DB 4;	Length 713;	
	Best Local Similarity	22.1%;	Pred. No. 0.36;			
	Matches	78;	Conservative	38;	Mismatches	157;
					Indels	80;
					Gaps	12
QY	30	LRPKLOQTTEETENACGRPISGVGVTVIVANGPKSGFGRLSCCKGHWICPCAGKVGAH	89			
	:	:	:	:	:	:
DB	302	MRVDVSQARTAAAMCTVALLLATPAATCAVVPALGARGLGSRAALAVAGSLNNKG--	359			
QY	90	RADLSQVVAHQLTGTS-----VAMVTMTMHTTGGQRLLDMLTGISAAMKAATNGR	140			
	:	:	:	:	:	:
DB	360	-TDILVALVGVELGLLSNEAYTMVAVALVTV---TASPALL-IWLEKRA---PPTQE	410			
QY	141	KWRTEREMGYCDGYRAVELTHGKNKGWEHVHALIMESGDVSENITLESFSDMPRWTSK	200			

```

Db      411 SARLEREEAARRAYIPGVE-----RILVIVAHALPGFATDIVESIVAS----- 454
Qy      201 LVSLGFAPLRNSGGLDVKRIKIGEGADQVLAAYLTIKIASGVMEVSGDGKSGRHGNR--- 257
Db      455 -----KRIKIGETVD-ITELSVGOAPSPRAAGEARSGIARLGARLRV 496
Qy      258 APWEIADVAVGDDPOLAELMREFFSGMGRRAIAMSRLRARGIGAEITDPAQIVEQES 317
Db      497 GIWRQRRELRSIGAILRASRDHDLVIGARSPARAGMSF-----GRLODAIVQRAESN 551
Qy      318 APVWVAIIPARSGMMIRTCAPYVGEIIGLVEAGATWENLRDHLHYRLPADV 370
Db      552 VLVVVGDPPEAAE---RASARRILVPIIG-----LEYSFAAADL 586

```

```

RESULT 13
US-09-568-486-11
; Sequence 11, Application US/09568486
; Patent No. 6355459
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goelzsch, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/568,486
; CURRENT FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 11
; LENGTH: 713
; TYPE: PRT
; ORGANISM: Sorangium cellulosum
US-09-568-486-11

```

```

Query Match      4.8%; Score 97; DB 4; Length 713;
Best Local Similarity 22.1%; Pred. No. 0.36;
Matches 78; Conservative 38; Mismatches 157; Indels 80; Gaps 12;

Qy      30 LRPKIQITSETFNACGRPIGSGVGVITVNGPKSGGGLRSCGKGCPCGAKVGAH 89
Db      302 MRVDVSQLRTPAAGVTALLATATPAKVPALGRLGIGSEBAALVAVGLNMKG-- 359
Qy      90 RADEISQVVAHQLTGTS-----VAMVTMTMRHTAGRLHDLMTGLSAAKKAATNGR 140
Db      360 -TDLVAIVGVELGILSNAYTMVAVALTV-----TASPALL-ITLEKRA---PPTQEE 410
Qy      141 RWRTEREMYGCDGYRAVEITHGKNGMHVHALLMFSGDVSENILSFSDAMPFRWTSK 200
Db      411 SARLEREEAARRAYIPGVE-----RILVIVAHALPGFATDIVESIVAS----- 454
Qy      201 LVSLGFAPLRNSGGLDVKRIKIGEGADQVLAAYLTIKIASGVMEVSGDGKSGRHGNR--- 257
Db      455 -----KRIKIGETVD-ITELSVGOAPSPRAAGEARSGIARLGARLRV 496
Qy      258 APWEIADVAVGDDPOLAELMREFFSGMGRRAIAMSRLRARGIGAEITDPAQIVEQES 317
Db      497 GIWRQRRELRSIGAILRASRDHDLVIGARSPARAGMSF-----GRLODAIVQRAESN 551
Qy      318 APVWVAIIPARSGMMIRTCAPYVGEIIGLVEAGATWENLRDHLHYRLPADV 370
Db      552 VLVVVGDPPEAAE---RASARRILVPIIG-----LEYSFAAADL 586

```

```

RESULT 14
US-09-568-472-11
; Sequence 11, Application US/09568472

```

```

; Patent No. 6358719
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goelzsch, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/568,472
; CURRENT FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 11
; LENGTH: 713
; TYPE: PRT
; ORGANISM: Sorangium cellulosum
US-09-568-472-11

```

```

Query Match      4.8%; Score 97; DB 4; Length 713;
Best Local Similarity 22.1%; Pred. No. 0.36;
Matches 78; Conservative 38; Mismatches 157; Indels 80; Gaps 12;

Qy      30 LRPKIQITSETFNACGRPIGSGVGVITVNGPKSGGGLRSCGKGCPCGAKVGAH 89
Db      302 MRVDVSQLRTPAAGVTALLATATPAKVPALGRLGIGSEBAALVAVGLNMKG-- 359
Qy      90 RADEISQVVAHQLTGTS-----VAMVTMTMRHTAGRLHDLMTGLSAAKKAATNGR 140
Db      360 -TDLVAIVGVELGILSNAYTMVAVALTV-----TASPALL-ITLEKRA---PPTQEE 410
Qy      141 RWRTEREMYGCDGYRAVEITHGKNGMHVHALLMFSGDVSENILSFSDAMPFRWTSK 200
Db      411 SARLEREEAARRAYIPGVE-----RILVIVAHALPGFATDIVESIVAS----- 454
Qy      201 LVSLGFAPLRNSGGLDVKRIKIGEGADQVLAAYLTIKIASGVMEVSGDGKSGRHGNR--- 257
Db      455 -----KRIKIGETVD-ITELSVGOAPSPRAAGEARSGIARLGARLRV 496
Qy      258 APWEIADVAVGDDPOLAELMREFFSGMGRRAIAMSRLRARGIGAEITDPAQIVEQES 317
Db      497 GIWRQRRELRSIGAILRASRDHDLVIGARSPARAGMSF-----GRLODAIVQRAESN 551
Qy      318 APVWVAIIPARSGMMIRTCAPYVGEIIGLVEAGATWENLRDHLHYRLPADV 370
Db      552 VLVVVGDPPEAAE---RASARRILVPIIG-----LEYSFAAADL 586

```

```

RESULT 15
US-09-567-899-11
; Sequence 11, Application US/09567899
; Patent No. 6383787
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goelzsch, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/567,899
; CURRENT FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 11
; LENGTH: 713
; TYPE: PRT

```


GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 8, 2003, 23:02:33 ; Search time 232 Seconds
(without alignments)
280.573 Million cell updates/sec

Title: US-10-007-527A-2

Perfect score: 2005
Sequence: 1 MTSVSAEHLSGKDRPPVIVS.....HLHTRLPADYRPPIISVRK 379

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 644079 seqs, 171749292 residues

Total number of hits satisfying chosen parameters: 644079

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications AA:*

```
1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/2/pubpaa/PC1_NEW_PUB.pep:*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/2/pubpaa/PC105_PUBCOMB.pep:*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep:*
10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep:*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep:*
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep:*
13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep:*
14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep:*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep:*
16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep:*
17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep:*
18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2005	100.0	379	US-10-007-527A-2	Sequence 2, Appli
2	2005	100.0	379	US-10-007-452-2	Sequence 2, Appli
3	609.5	30.4	459	US-10-007-527A-21	Sequence 21, Appli
4	609.5	30.4	459	US-10-007-452-21	Sequence 21, Appli
5	326	16.3	427	US-09-835-381-6	Sequence 6, Appli
6	322	16.1	451	US-10-007-527A-24	Sequence 24, Appli
7	322	16.1	451	US-10-007-452-24	Sequence 24, Appli
8	318	15.9	427	US-09-835-381-8	Sequence 8, Appli
9	318	15.9	427	US-10-196-232-7	Sequence 7, Appli
10	316	15.8	456	US-10-007-527A-22	Sequence 22, Appli
11	316	15.8	456	US-10-007-452-22	Sequence 22, Appli
12	282	14.1	466	US-09-826-191-9	Sequence 9, Appli
13	282	14.1	466	US-10-263-666-9	Sequence 9, Appli
14	282	14.1	466	US-10-261-481-9	Sequence 9, Appli
15	282	14.1	466	US-10-261-942-9	Sequence 9, Appli

16	266.5	13.3	528	US-10-007-527A-23	Sequence 23, Appli
17	266.5	13.3	528	US-10-007-452-23	Sequence 23, Appli
18	103.5	5.2	417	US-10-156-761-10798	Sequence 10798, A
19	103	5.1	647	US-10-021-660-94	Sequence 94, Appli
20	100.5	5.0	544	US-10-156-761-8540	Sequence 8540, Ap
21	97	4.8	713	US-10-014-717-11	Sequence 11, Appli
22	96	4.8	7746	US-10-156-761-7965	Sequence 7965, Ap
23	95.5	4.8	3798	US-10-014-717-6	Sequence 6, Appli
24	95.5	4.8	4471	US-10-205-032-10	Sequence 10, Appli
25	94.5	4.7	877	US-10-156-761-13758	Sequence 13758, A
26	94	4.7	6146	US-10-156-761-10436	Sequence 10436, A
27	93.5	4.7	356	US-09-764-853-570	Sequence 570, App
28	92	4.6	310	US-10-156-761-9467	Sequence 9467, Ap
29	91.5	4.6	981	US-10-156-761-15071	Sequence 15071, A
30	91.5	4.6	5245	US-10-329-079-11	Sequence 11, Appli
31	91	4.5	245	US-10-156-761-14296	Sequence 14296, A
32	91	4.5	956	US-09-815-242-11925	Sequence 11925, A
33	90.5	4.5	309	US-10-306-762-88	Sequence 88, Appli
34	89.5	4.5	585	US-09-728-626-6032	Sequence 6032, Ap
35	89.5	4.5	3352	US-10-156-761-7961	Sequence 7961, Ap
36	89	4.4	3739	US-09-861-289-33	Sequence 33, Appli
37	89	4.4	3739	US-09-860-846-33	Sequence 33, Appli
38	89	4.4	3739	US-09-988-3848-33	Sequence 33, Appli
39	89	4.4	3739	US-09-836-821-33	Sequence 33, Appli
40	89	4.4	3739	US-09-793-708-2	Sequence 2, Appli
41	89	4.4	3739	US-10-201-765-3	Sequence 3, Appli
42	89	4.4	3739	US-10-160-539-2	Sequence 2, Appli
43	89	4.4	3739	US-10-271-889-33	Sequence 33, Appli
44	89	4.4	6291	US-10-329-079-41	Sequence 41, Appli
45	89	4.4	11877	US-09-861-289-6	Sequence 6, Appli

ALIGNMENTS

```
RESULT 1
US-10-007-527A-2
; Sequence 2, Application US/10007527A
; Publication No. US20030044807A1
; GENERAL INFORMATION:
; APPLICANT: Tomb, Jean-Francois
; APPLICANT: Bremucci, Michael G.
; APPLICANT: Cheng, Qiong
; APPLICANT: Koestelchka, Kristy N.
; TITLE OF INVENTION: Rhodococcus Cloning and Expression Vectors
; FILE REFERENCE: C11709 US NA
; CURRENT APPLICATION NUMBER: US/10/007,527A
; PRIOR FILING DATE: 2001-12-05
; PRIOR APPLICATION NUMBER: 60/254,868
; PRIOR FILING DATE: 2000-12-12
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 2
; LENGTH: 379
; TYPE: PRT
; ORGANISM: Rhodococcus AN12
US-10-007-527A-2
```

Query Match	Score	DB	Length	Matches	Score	DB	Length	Matches	Score	DB	Length	Matches
Best Local Similarity	100.0%			Pred. No. 5.3e-195;				Indels	0;			Gaps
Matches	379;	Conservative	0;	Mismatches	0;			Indels	0;			Gaps
1	MTSVSAEHLSGKDRPPVIVS	1	MTSVSAEHLSGKDRPPVIVS	60								
1	MTSVSAEHLSGKDRPPVIVS	1	MTSVSAEHLSGKDRPPVIVS	60								
61	GPKSGFGGLRSCKGKMTIC	61	GPKSGFGGLRSCKGKMTIC	120								
61	GPKSGFGGLRSCKGKMTIC	61	GPKSGFGGLRSCKGKMTIC	120								
121	RLHDMTGLSAAMKATNGR	121	RLHDMTGLSAAMKATNGR	180								
121	RLHDMTGLSAAMKATNGR	121	RLHDMTGLSAAMKATNGR	180								

```
QY 181 VSENILSFSDAMPDRMTSKVLISGFAAPLRNSGGLDVKRIKIGEGADQVLAAYLTXTIASGV 240
DB 181 VSENILSFSDAMPDRMTSKVLISGFAAPLRNSGGLDVKRIKIGEGADQVLAAYLTXTIASGV 240
QY 241 GMEVSGDGKSGRHNRAPEIADVAGDPOALELMEFEFGSGRRALIAWSRGLRARA 300
DB 241 GMEVSGDGKSGRHNRAPEIADVAGDPOALELMEFEFGSGRRALIAWSRGLRARA 300
QY 301 GIGAEITDAQIVQEESAPVMAIIPASMMWIRTCAPYVFEIIGLVEAGATWENLRDH 360
DB 301 GIGAEITDAQIVQEESAPVMAIIPASMMWIRTCAPYVFEIIGLVEAGATWENLRDH 360
QY 361 LHYRLPADVPRPPIISVRK 379
DB 361 LHYRLPADVPRPPIISVRK 379
```

```
RESULT 2
US-10-007-452-2
; Sequence 2, Application US/10007452
; Publication No. US20030093701A1
; GENERAL INFORMATION:
; APPLICANT: Tomb, Jean-Francois
; APPLICANT: Bramucci, Michael G.
; APPLICANT: Cheng, Qiong
; APPLICANT: Kostichka, Kristy N.
; TITLE OF INVENTION: Rhodococcus Cloning and Expression Vectors
; FILE REFERENCE: CL1709 US NA
; CURRENT APPLICATION NUMBER: US/10/007,452
; PRIOR FILING DATE: 2001-11-08
; PRIOR APPLICATION NUMBER: 60/254,868
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 2
; LENGTH: 379
; TYPE: PRF
; ORGANISM: Rhodococcus AN12
US-10-007-452-2
```

```
Query Match 100.0%; Score 2005; DB 15; Length 379;
Best Local Similarity 100.0%; Pred. No. 5,3e-195;
Matches 379; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MTSVSAEHLISGKDRPPVLYSSDKRGIRHRLRPKLOITTSSTFNACGRPIISGVNGVTYN 60
DB 1 MTSVSAEHLISGKDRPPVLYSSDKRGIRHRLRPKLOITTSSTFNACGRPIISGVNGVTYN 60
QY 61 GPKSGFGGIRSCGKGMICPCCAKVGARADEISQVVAHQLGTSVAVMTMTMTHTAQ 120
DB 61 GPKSGFGGIRSCGKGMICPCCAKVGARADEISQVVAHQLGTSVAVMTMTMTHTAQ 120
QY 121 RLHDLMTGLSAANKAATNGRWRTEREMTCDDGYRAVEITTHGKNGMHVHVALIMFSGD 180
DB 121 RLHDLMTGLSAANKAATNGRWRTEREMTCDDGYRAVEITTHGKNGMHVHVALIMFSGD 180
QY 181 VSENILSFSDAMPDRMTSKVLISGFAAPLRNSGGLDVKRIKIGEGADQVLAAYLTXTIASGV 240
DB 181 VSENILSFSDAMPDRMTSKVLISGFAAPLRNSGGLDVKRIKIGEGADQVLAAYLTXTIASGV 240
QY 241 GMEVSGDGKSGRHNRAPEIADVAGDPOALELMEFEFGSGRRALIAWSRGLRARA 300
DB 241 GMEVSGDGKSGRHNRAPEIADVAGDPOALELMEFEFGSGRRALIAWSRGLRARA 300
QY 301 GIGAEITDAQIVQEESAPVMAIIPASMMWIRTCAPYVFEIIGLVEAGATWENLRDH 360
DB 301 GIGAEITDAQIVQEESAPVMAIIPASMMWIRTCAPYVFEIIGLVEAGATWENLRDH 360
QY 361 LHYRLPADVPRPPIISVRK 379
DB 361 LHYRLPADVPRPPIISVRK 379
```

```
RESULT 3
US-10-007-527A-21
; Sequence 21, Application US/10007527A
; Publication No. US20030044807A1
; GENERAL INFORMATION:
; APPLICANT: Tomb, Jean-Francois
; APPLICANT: Bramucci, Michael G.
; APPLICANT: Cheng, Qiong
; APPLICANT: Kostichka, Kristy N.
; TITLE OF INVENTION: Rhodococcus Cloning and Expression Vectors
; FILE REFERENCE: CL1709 US NA
; CURRENT APPLICATION NUMBER: US/10/007,527A
; PRIOR FILING DATE: 2001-12-05
; PRIOR APPLICATION NUMBER: 60/254,868
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 21
; LENGTH: 459
; TYPE: PRF
; ORGANISM: Arcanobacterium pyogenes
US-10-007-527A-21
```

```
Query Match 30.4%; Score 609.5; DB 15; Length 459;
Best Local Similarity 36.3%; Pred. No. 3.8e-53;
Matches 148; Conservative 62; Mismatches 133; Indels 65; Gaps 15;

QY 5 SAELHSGKDRPPVLYSSD-----KRGIRHRLRPKLOITTSSTFNACGRPIISGVNGV 56
DB 67 SAHPLGNTVTLFPVNSKSTKTAERSERYELDGLMELTISVRCGVPAPLVSL 126
QY 57 TVVNGPKSGFGGIRSCGKGMICPCCAKVGARADEISQVVAHQLGTSVAVMTMTMT 115
DB 127 RAKSDGKAGVGGHGTGOSVACVCSAKIABRTDLDQVVDHAVKGMVSMTLTQR 186
QY 116 HTAGRLHDLMTGLSAANKAATNGRWRTEREMTCDDGYRAVEITTHGKNGMHVHVALI 175
DB 187 HHKGQGLHMDALSTANRWYTSGRWRTEFKEQFGLVGYRAVEITTHGKNGMHVHVALI 246
QY 176 MESGDVSENILSF-----SDAMPDRMTSKVLISGFAAPLRNSGG 215
DB 247 ISEKD--PLVSTFYQKQGRRLPYPEIYMSDFTAEKWEAGLAHGVDF-LRDSGG 302
QY 216 L-----DVKRIKIGEGADQVLAAYLTKI--ASGVMEVSGDGKSGRHNRAPEIADVA 266
DB 303 LDWYAKXARALGN-----YISKMGTSIDAISSEVTLGSKFKAARNGRTTPQILADI 354
QY 267 VG-GDPQALELMEFEFGSGRRALIAWSRGLRARAQIGAEITDAQIVQEESAPVMAI 325
DB 355 LSLGVDVLDLKLMEYEEKASFGRRALTWKGLRDMANLGVQSDDEIA--SEETIGDEALALF 413
QY 326 PARSMWMTIRTCAPYVFG--EILGLVEAGA-----TWENLRDLHRLP 366
DB 414 THDAWRQYR-----FGAEILDVYESGRRAAAYRWLDFR--TDWSLP 455
```

```
RESULT 4
US-10-007-452-21
; Sequence 21, Application US/10007452
; Publication No. US20030093701A1
; GENERAL INFORMATION:
; APPLICANT: Tomb, Jean-Francois
; APPLICANT: Bramucci, Michael G.
; APPLICANT: Cheng, Qiong
; APPLICANT: Kostichka, Kristy N.
; TITLE OF INVENTION: Rhodococcus Cloning and Expression Vectors
; FILE REFERENCE: CL1709 US NA
; CURRENT APPLICATION NUMBER: US/10/007,452
; PRIOR FILING DATE: 2001-11-08
; PRIOR APPLICATION NUMBER: 60/254,868
; NUMBER OF SEQ ID NOS: 30
```

SOFTWARE: Microsoft Office 97
SEQ ID NO 21
LENGTH: 459
TYPE: PRF
ORGANISM: Arcanobacterium pyogenes
US-10-007-452-21

Query Match 30.4%; Score 609.5; DB 15; Length 459;
Best Local Similarity 36.3%; Pred. No. 3.8e-53;
Matches 148; Conservative 62; Mismatches 133; Indels 65; Gaps 15;

QY 5 SAEHLSGKDRPPVLYSSD-----KRGIRHRLRKLQOITTSFPMNAGR-PISGVAGV 56
DB 67 SATHPLGNTVLTFTFPVSNESKTKTAKSRSERYELRGLAEISTIESVRKCGRPVAPLVSL 126
QY 57 TVNPPKSGSGFGGLSCGKMWICPCGAGVGAHRADEISQVVAHOLGTG-SVAMTMTMR 115
DB 127 RAKSGGKAGYGGGLTTCGWAACPVCSAKIAPRKTDLQGVDDHAVKGMTSMTLTQK 186
QY 116 HTAGORLHDLMTGLSAAMKAAATNGRRMTEREMYGCDGVRAVEITTHGKNGMHVHALL 175
DB 187 HHKGGGLKHLMDALSTAMNRVTSGRRWTFEKGQFGLVGVYRANEITHGKGMHVHSHVLI 246
QY 176 MESGVSENIIESF-----SDAMPRTSKLVSLGFAAPLRNSGG 215
DB 247 ISEKD---PLTSTFYQKQKGRRLPYPEIYMSDFIAEFWEAGAKHGVDF-LRDSGG 302
QY 216 L-----DYRKIGGEADQVLAAYLTKE--ASGVGMEVSGDGKSGRGNRAPWEIAYDA 266
DB 303 LMTVYAKDKRAIGN-----YVKMQCTSTAISEVTLGGFKAKRNGNRKTFPQLADI 354
QY 267 VG-GDPQALMLWREFEFGSGMRRALAMSRGLRARAAGLGAELTDQIVQEESAPVWALI 325
DB 355 LSLGVDVDDKLMKEVEKSKFGRALTWKGLDMANLVGEQGDDEILA-SEIIGDEALALF 413
QY 326 PARSMNTITTCAPYVFG--EITGLVEAGA-----TWENTRIDLIHRLP 366
DB 414 THDAWRQYVR-----FGAAELLDVYESGGRAAAYRWLDERE-IDWSLP 455

RESULT 5
US-09-835-381-6
Sequence 6, Application US/09835381
Patent No. US20020045223A1
GENERAL INFORMATION:
APPLICANT: SUGA, Mikiko
APPLICANT: ASAKURA, Yoko
APPLICANT: MORI, Yukiko
APPLICANT: KURAHASHI, Osamu
TITLE OF INVENTION: ARGININE REPRESSOR DEFICIENT STRAIN OF CORYNEFORM BACTERIUM AND N
FILE REFERENCE: 206018USO
CURRENT APPLICATION NUMBER: US/09/835,381
CURRENT FILING DATE: 2001-04-17
PRIOR APPLICATION NUMBER: JP2000-129167
PRIOR FILING DATE: 2000-04-28
NUMBER OF SEQ ID NOS: 22
SOFTWARE: PatentIn version 3.1
SEQ ID NO 6
LENGTH: 427
TYPE: PRF
ORGANISM: Brevibacterium lactofermentum
US-09-835-381-6

Query Match 16.3%; Score 326; DB 9; Length 427;
Best Local Similarity 29.1%; Pred. No. 2.2e-24;
Matches 98; Conservative 58; Mismatches 131; Indels 50; Gaps 15;
QY 34 LQOITTSFPMNAGR-----PISGVNQTIVNGKSGFGGLRSCGKMWICPCGAGV 86
DB 27 MYKINSKALAGCHWRDEAVAVSWSS-----NG-ASQFGLQNSHRSWSPLALELV 79

QY 87 GAHRADEISQVVAHOLGTGSVAMVTM-TMRHTAGORLHDLMTGLSAAMKAAATNGRRMTE 145
DB 80 MGEERIELAIATYKHLAAGALMMFVGTVRRNSQSFQAQVAGIKTAISSVWKTSGMKE 139
QY 146 REMYCDGYVRAVEITHG-KNGMHVHVALIMFSGSENILIESFSDDAMPRTSKLVSL 204
DB 140 PARYGVHTTSDYEVTDTSKMGMLHHRMLFLFLRPLSDDLKAFEDSMFSRMSAGVYKA 199
QY 205 GPAAPLRNSG-GLD-VRKIGGEADQVLAAYLTKIASGVGMEVSGDGKSGRGNRAPWEI 262
DB 200 GMDAPLRHGVKLDQVSTWVGDAK-MATYLAK---GMSQELTGSATKATASKGSYTPQM 255
QY 263 AYDAVGDPQALF-----LWREFEFGSGMRRALAMSRGLRARAAGLGAELTDQI--- 311
DB 256 -LDMQLADQSDAGEMDAVLVARMBREYEVSGKXLR-SWSRG-AKRALIDYIDAVBRE 311
QY 312 VEOE-----ESAPVWALIIPARSMNTIT 335
DB 312 MEEELIYKLAGLEAPERVASTVVAVALVAPDDMKLIQS 348

RESULT 6
US-10-007-527A-24
Sequence 24, Application US/10007527A
Publication No. US20030044807A1
GENERAL INFORMATION:
APPLICANT: Tomb, Jean-Francois
APPLICANT: Bramucci, Michael G.
APPLICANT: Cheng, Qiong
APPLICANT: Kostichka, Kristy N.
TITLE OF INVENTION: Rhodococcus Cloning and Expression Vectors
FILE REFERENCE: CL1709 US NA
CURRENT APPLICATION NUMBER: US/10/007,527A
CURRENT FILING DATE: 2001-12-05
PRIOR APPLICATION NUMBER: 60/254,868
PRIOR FILING DATE: 2000-12-12
NUMBER OF SEQ ID NOS: 30
SOFTWARE: Microsoft Office 97
SEQ ID NO 24
LENGTH: 451
TYPE: PRF
ORGANISM: Streptomyces nigritifaciens
US-10-007-527A-24

Query Match 16.1%; Score 322; DB 15; Length 451;
Best Local Similarity 30.8%; Pred. No. 6e-24;
Matches 101; Conservative 41; Mismatches 110; Indels 76; Gaps 15;

QY 49 PISGV-----NGVTIVNGPKSGFGGLRSCGKMWICPCGAGVGAHRADEISQVVAHOL 102
DB 3 PASGVIAQTAAGTSVVL-----GLMRCGRIMLCPVCAATIBHKRAEITAAVWEI 54
QY 103 GTGSAVA-WYTMTRHTAGORLHDLMTGLSAAMKAAATNGRR-----WRTS----- 145
DB 55 KRGTAYLVTFARAGHTDLRLMDALQGTREKTDAPRRPQAYORLTGTGWAGRRAYD 114
QY 146 -----REMYCDGYVRAVEITHG-KNGMHVHVALIMFSGDV-----SENILIESF 189
DB 115 GHRADRBGIRDRIGYVGMIRATVTVQINGMHPIHALVVGRTBBERAKQVGVF 174
QY 190 --SDAMPRTSKLVSLGFAAPLR-----NSGGLDVKRIIGEAD-QVLAAYITKTA 237
DB 175 BPSEALDWMQOMPAV-WTAAIRKVNQFTPDDBRGVDFKLETERDANDLAETAKQ 233
QY 238 SG--VGMVEVSGDGSGRGNRAPWEI---AYDAVG---DQA-----LELWREFE 281
DB 234 DKAAPALREARADLKTANGVNAPELLGRIGDLTGCTEDDAAGVSGILEMMLARWHEE 293
QY 282 FSGMGRRAIAMSRLRARAAGLGAELTDA 309
DB 294 RATKGRRALEWTRYLRLQMLGDGDTEA 321

	Query Match	16.1%	Score 322;	DB 15,	length 451;	
	Best Local Similarity	30.8%	Pred. No. 6e-24;			
	Matches 101; Conservative	41;	Mismatches 110;	Indels 76;	Gaps 15	
QY	49 PISGV-----NGVTTVNGPKSGFGSLSCGKGNICCCAGKVGARPADEISQVAHQI	102				
Db	3 PASGYIAQTAAAGTSVL-----GLMRCGRIMLCPVCATIRHKKAEETIAAVEMVI	54				
QY	103 GTGSVA-WVTMTMETTAQQLRLDMLTGTLSAANKATINRR-----WRE----	145				
Db	55 KRGGYAYLVPTTARHGHTDRDLADLMDAOIGTRKTDAARRPGAYRLLITGGTWAGRAXD	114				
QY	146 -----REMWCDDGVRAVEITETHCK-NGMHVHYALMEFSDV-----SENILSEF	189				
Db	115 GHRAADREGIDRIDRGIVEMIRATEVTAVGOINSMHPHIHAIVLVGRTGEGBERAKOIVGITF	174				
QY	190 --SDAMPFRWTSKLVSLSGFAAPLR-----NSGGLDVVKXIGEAD-QVLAAYLTIXIA	237				
Db	175 KPSEAALDEMGQWMAV-WTALAARKVNIPQFTPDDRHGVDFKRLETERDANDLAIFYAKTO	233				
QY	238 SG-VGMVEVSGDGKSGRHGNFAPMEI--AVDAYGG--DPQA-----LELMREE	281				
Db	234 DGAFALETAFRADLKTXANGGANVAPELLIGRIDTLGXTMEDDPAAGVSLLENNIARWHEYE	293				
QY	282 FGSMGRRAIAMSGRARABAGLAEILTDA	309				
Db	294 RATKGRRAIEMTRYLRQMILGDGGDTAA	321				

```

Query Match      15.9%; Score 318; DB 9; Length 427;
Best Local Similarity 28.8%; Pred. No. 1,4e-23;
Matches 97; Conservative 58; Mismatches 132; Indels 50; Gaps 15;

QY      34 LQOITSETFNACGR-----PISGVNGVITVNEPKGSGFGGLRGCGKWTCPCCAGKY 86
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      27 MYKITSKALAGCHRRRDEAVAVSS-----NG--AAGFQGLQNSHSRMSSLALELV 79
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY      87 GAHRDAEISQVVAHQJGTSVAVMT--TNRHTAGQRLHDMTGLSLAAKAAITGRWRKE 145
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      80 MGEERIELATATKNHLLAAGALMPFGTAVRHNSQSFAQVEAGIKTAYSSMVTYSOMKKE 139
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY      146 RENVGCDGYRAVEITHG--KNGHNVVHALLMSSGVSENILSEFSDAGNDRPTSKLVL 204
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      140 RARGVHTYSDEYEVDTSWANGWHLHRNMILLFDRPLISDDLELAFEDSMFSRWAGVTA 199
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY      205 GFAAPLENSG-GID-VKLTIGEADQVLAAYLTKTASGVMEVSGDGKSGRHNRAPEWI 262
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      200 GMDAPLRHEHVKLDQVSTWGGDAK--MATYIAK---GMSGELLSGATKTATSKSGSYNPFQM 255
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY      263 AVDAVGDGPALR-----LWREFPGSGERRAIAWSFGLPARAGLCAELTDAQI--- 311
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      256 -LDMTADQSDAGEMDAVLARVRREYEVGSKNLR--SWSRG--AKRALGIDYIDADVRE 311
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY      312 VEQR-----ESAPVVAIIIPASMMIIR 335
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      312 MEELVYKLAGLEADERVESTRVAVALKKPDMDKLIQS 348
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

	Query Match	15.9%	Score 318;	DB 15;	Length 427;
	Best Local Similarity	28.8%	Pred. No. 1.4e-23;		
	Matches	97;	Conservative	58;	Mismatches 133; Indels 50; Gaps 15;
QY	34	LQQTITSETFNACGR-----PISGVGVTIVNGPKSGFGILRSCKGMITCECCAKV	86		
		: : : : :	:	:	:
Dd	27	MYKINSPALAGCHWRDRDEAVAVSWSS-----NG--ASQEGQLNHSRFGSSLAELEV	79		
		: : : : :	:	:	:
QY	87	GAHRDELISQVVAHDLGTGSVAATYM-TMRTTAGCRLLDMLTGSLAAKATATNRBMTTE	145		
		: : : : :	:	:	:
Dd	80	MGERRIELAIATKNLILAAGALLMFVFGRVRRNRSSFAOVAIGIKTAIVSSVKTSQMKEE	139		
		: : : : :	:	:	:
QY	146	REMYCDDGVRAVEITTHG-KMGHHVHYALLFMFSDVBENLTIESPSDAMPDRMTSKVSL	204		
		: : : : :	:	:	:
Dd	140	RARYGEVHTYSLEYTDSANAMHLHRMLFLFDLPJSDDEKLKAFEDSMFMRWSAGVYKA	199		
		: : : : :	:	:	:
QY	205	GFAAPLRNSG-GLD-VRKTIIGEADOVTLAAYLTKLASGVMEVSGDGKSGKHGNRAPWEI	262		

```

Db      200 GMDAPLREHGVLTQVSTMGDAK--MATYLAK---GMSGLTSATKASGSTYPPQM 255
      263 AVDAVGDDPALE-----LMREFEGSGRRALAMSGILPARAGAEITDAOI--- 311
      256 -IDMLADQSDAGBDMVAVLVARREYEVGSKULRS--SMERG--AKRALGIDYIDAVRE 311
Qy      312 VEOE-----ESAPVVAIIPARSMIMRT 335
      312 MEEELYKLAGLEAPERVESTRAVALVKPDWKLIQS 348

```

RESULT 10
US-10-007-527a-22

```

; Sequence 22, Application US/10007527A
; Publication No. US20030044807A1
; GENERAL INFORMATION:
; APPLICANT: Tomb, Jean-Francois
; APPLICANT: Bramucci, Michael G.
; APPLICANT: Cheng, Qiong
; APPLICANT: Kostichka, Kristy N.
; TITLE OF INVENTION: Rhodococcus Cloning and Expression Vectors
; FILE REFERENCE: C11709 US NA
; CURRENT APPLICATION NUMBER: US/10/007,527A
; PRIOR FILING DATE: 2001-12-05
; PRIOR APPLICATION NUMBER: 60/254,868
; PRIOR FILING DATE: 2000-12-12
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 22
; LENGTH: 456
; TYPE: PRT
; ORGANISM: Streptomyces lividans
US-10-007-527a-22

```

Query Match 15.8%; Score 316; DB 15; Length 456;
Best Local Similarity 30.5%; Pred. No. 2.5e-23;
Matches 100; Conservative 40; Mismatches 112; Indels 76; Gaps 15;

```

Qy      49 PISGV-----NGVTIVGPKSGSGFGLRSCGKGMICPCGAGKVGARADEISQVAAHQL 102
      3 PASGVIYAQTAAAGTSVVL-----GLMRCGRIMLCPVCAATIRHRAEITAAVEMI 54
Db      103 GTGSVA-WVTMTMRHTAGORLHDLMTGLSAWKAATNGRR-----WRTE----- 145
      55 KKGTAIYVTFARHGHTRDLADLMDALQSTRKTPDSPRRPGAYQRLITGTMAGRRAD 114
      146 -----REMYGCDGVRAVEITTHGK--NGMHVYHALMFSGDV-----SENTLESE 189
      115 GHRADREGIRIDRIGYVGMIRATEVVGQINGMHPHIAVLVGGRTGERSAKQIVATF 174
Qy      190 --SDAMFDRWTSKLVLGFAAPLR-----NSGLDVRKIGIGAD--QVLAAYLTXTA 237
      175 EPTGALDEWQGMRSV-WTAAIKXNPAFTPDGRGVDFKLETERDANDLAEYIAKQ 233
Db      238 SG--VGMVEGSGDGKSGRHNRAPEI--AVDAVG--DPOA-----LELMREFE 281
      234 DGKAPALBELARADLKTATGCVNAPFELIGRIDLTGQMTEDDAAGVSGLEMLSRMHEYE 293
Qy      282 FSGMGRRAIAMSRLPARAGAEITDA 309
      294 RATRGRALEMTRYLROMLGLDGDTEA 321
Db

```

RESULT 11
US-10-007-452-22

```

; Sequence 22, Application US/10007452
; Publication No. US20030093701A1
; GENERAL INFORMATION:
; APPLICANT: Tomb, Jean-Francois
; APPLICANT: Bramucci, Michael G.
; APPLICANT: Cheng, Qiong
; APPLICANT: Kostichka, Kristy N.

```

```

; TITLE OF INVENTION: Rhodococcus Cloning and Expression Vectors
; FILE REFERENCE: C11709 US NA
; CURRENT APPLICATION NUMBER: US/10/007,452
; PRIOR FILING DATE: 2001-11-08
; PRIOR APPLICATION NUMBER: 60/254,868
; PRIOR FILING DATE: 2000-12-12
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 22
; LENGTH: 456
; TYPE: PRT
; ORGANISM: Streptomyces lividans
US-10-007-452-22

```

Query Match 15.8%; Score 282; DB 9; Length 466;
Best Local Similarity 30.0%; Pred. No. 7.2e-20;
Matches 87; Conservative 40; Mismatches 119; Indels 44; Gaps 15;

```

Qy      49 PISGV-----NGVTIVGPKSGSGFGLRSCGKGMICPCGAGKVGARADEISQVAAHQL 102
      3 PASGVIYAQTAAAGTSVVL-----GLMRCGRIMLCPVCAATIRHRAEITAAVEMI 54
Db      103 GTGSVA-WVTMTMRHTAGORLHDLMTGLSAWKAATNGRR-----WRTE----- 145
      55 KKGTAIYVTFARHGHTRDLADLMDALQSTRKTPDSPRRPGAYQRLITGTMAGRRAD 114
Qy      146 -----REMYGCDGVRAVEITTHGK--NGMHVYHALMFSGDV-----SENTLESE 189
      115 GHRADREGIRIDRIGYVGMIRATEVVGQINGMHPHIAVLVGGRTGERSAKQIVATF 174
Qy      190 --SDAMFDRWTSKLVLGFAAPLR-----NSGLDVRKIGIGAD--QVLAAYLTXTA 237
      175 EPTGALDEWQGMRSV-WTAAIKXNPAFTPDGRGVDFKLETERDANDLAEYIAKQ 233
Db      238 SG--VGMVEGSGDGKSGRHNRAPEI--AVDAVG--DPOA-----LELMREFE 281
      234 DGKAPALBELARADLKTATGCVNAPFELIGRIDLTGQMTEDDAAGVSGLEMLSRMHEYE 293
Qy      282 FSGMGRRAIAMSRLPARAGAEITDA 309
      294 RATRGRALEMTRYLROMLGLDGDTEA 321
Db

```

RESULT 12

```

US-09-826-191-9
; Sequence 9, Application US/09826191
; Patent No. US200200665A1
; GENERAL INFORMATION:
; APPLICANT: Stoddard, Steven F.
; APPLICANT: Ketogulonigenium
; TITLE OF INVENTION: KETOGLONIGENIUM ENDOGENOUS PLASMIDS
; FILE REFERENCE: 1533.1260001
; CURRENT APPLICATION NUMBER: US/09/826,191
; PRIOR FILING DATE: 2001-04-05
; PRIOR APPLICATION NUMBER: US 60/194,627
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: Patent version 3.0
; SEQ ID NO 9
; LENGTH: 466
; TYPE: PRT
; ORGANISM: Ketogulonigenium
US-09-826-191-9

```

Query Match 14.1%; Score 282; DB 9; Length 466;
Best Local Similarity 30.0%; Pred. No. 7.2e-20;
Matches 87; Conservative 40; Mismatches 119; Indels 44; Gaps 15;

```

Qy      60 NGPKSGFGLRSCGKGMICPCGAGKVGARADEISQVAAHQLGTGSV-AMVTMTMRHTA 118
      109 DGGRRAFFSGLAQCRNVWGCAVCSARIAQIRSENMHLLAWARDNGFVPLVTLTQHK 168
Qy      119 GQRHDLMTGLSAWKAATNGRRMTEREMYGCD---GYVAVEITTHG-KNGMHVYHVA 173

```

```
Db      169 GDSLFDDLQNMKKAQRLRORREWR-----DLPEVGSVTSTETHSYANGMHPHFE 220
Qy      174 L-LMPSGDVSE--NILESFSDFMFRMTSKLVSIGFAPLNNSGGLDVYKIGEGADQVLA 230
Db      221 IVLRAQDESEALHLMQRLGDA---WRACLKGYGMWG---NDAAFDVRGAANGD----- 269
Qy      231 AYLTKIASGVMEVSGDGKSGRHGNRAPMEIAVDVAGDPOALEMREFFGSGWG--RR 288
Db      270 -YVAK--WGAEEELTSSSKSGKRGRTPROLL--QAQDD---GLMIEFNATSGKRRR 320
Qy      289 AIAWSRGRLRARGLAGELTD---AQIVEQESAPVWVAIIPASMMMR 334
Db      321 QLVMSQGLKEECGL-VELDDDEAMAEVDAEAGSGPEIVAEMDNBSGKQVR 369
```

```
RESULT 13
US-10-263-666-9
; Sequence 9, Application US/10263666
; Publication No. US20030073224A1
; GENERAL INFORMATION:
; APPLICANT: D'Elia, John
; APPLICANT: Stoddard, Steven F.
; TITLE OF INVENTION: KETOGLONIGENIUM ENDOGENOUS PLASMIDS
; FILE REFERENCE: 1533.1260004
; CURRENT APPLICATION NUMBER: US/10/263,666
; PRIOR FILING DATE: 2002-10-04
; PRIOR FILING DATE: 2001-04-05
; PRIOR APPLICATION NUMBER: US 60/194,627
; PRIOR FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 9
; LENGTH: 466
; TYPE: PRT
; ORGANISM: Ketogulonigenium
US-10-263-666-9
```

```
Query Match      14.1%; Score 282; DB 15; Length 466;
Best Local Similarity 30.0%; Pred. No. 7.2e-20;
Matches 87; Conservative 40; Mismatches 119; Indels 44; Gaps 15;

Qy      60 NBPKSGFGGLRSCGKGMICPCGAGKVGARADEISQVVAHQLTGGSV-AMVTMTMRHTA 118
Db      109 DGGRRAFFSGLAQCRRNWGCVCASARIADIRSENMHLLAMARDNGFVPLITLTAQHKA 168
Qy      119 GQRLHDLMTGLSAAMKAAATNGRRMTEREMYGCD---GYRAVEITHG-KNGWHVHYHA 173
Db      169 GDSLFDDLQNMKKAQRLRORREWR-----DLPEVGSVTSTETHSYANGMHPHFE 220
Qy      174 L-LMPSGDVSE--NILESFSDFMFRMTSKLVSIGFAPLNNSGGLDVYKIGEGADQVLA 230
Db      221 IVLRAQDESEALHLMQRLGDA---WRACLKGYGMWG---NDAAFDVRGAANGD----- 269
Qy      231 AYLTKIASGVMEVSGDGKSGRHGNRAPMEIAVDVAGDPOALEMREFFGSGWG--RR 288
Db      270 -YVAK--WGAEEELTSSSKSGKRGRTPROLL--QAQDD---GLMIEFNATSGKRRR 320
Qy      289 AIAWSRGRLRARGLAGELTD---AQIVEQESAPVWVAIIPASMMMR 334
Db      321 QLVMSQGLKEECGL-VELDDDEAMAEVDAEAGSGPEIVAEMDNBSGKQVR 369
```

```
RESULT 14
US-10-261-481-9
; Sequence 9, Application US/10261481
; Publication No. US20030077830A1
; GENERAL INFORMATION:
; APPLICANT: D'Elia, John
; APPLICANT: Stoddard, Steven F.
; TITLE OF INVENTION: KETOGLONIGENIUM ENDOGENOUS PLASMIDS
; FILE REFERENCE: 1533.1260003
```

```
; CURRENT APPLICATION NUMBER: US/10/261,481
; CURRENT FILING DATE: 2002-10-02
; PRIOR APPLICATION NUMBER: US 09/826,191
; PRIOR FILING DATE: 2001-04-05
; PRIOR APPLICATION NUMBER: US 60/194,627
; PRIOR FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 9
; LENGTH: 466
; TYPE: PRT
; ORGANISM: Ketogulonigenium
US-10-261-481-9
```

```
Query Match      14.1%; Score 282; DB 15; Length 466;
Best Local Similarity 30.0%; Pred. No. 7.2e-20;
Matches 87; Conservative 40; Mismatches 119; Indels 44; Gaps 15;

Qy      60 NBPKSGFGGLRSCGKGMICPCGAGKVGARADEISQVVAHQLTGGSV-AMVTMTMRHTA 118
Db      109 DGGRRAFFSGLAQCRRNWGCVCASARIADIRSENMHLLAMARDNGFVPLITLTAQHKA 168
Qy      119 GQRLHDLMTGLSAAMKAAATNGRRMTEREMYGCD---GYRAVEITHG-KNGWHVHYHA 173
Db      169 GDSLFDDLQNMKKAQRLRORREWR-----DLPEVGSVTSTETHSYANGMHPHFE 220
Qy      174 L-LMPSGDVSE--NILESFSDFMFRMTSKLVSIGFAPLNNSGGLDVYKIGEGADQVLA 230
Db      221 IVLRAQDESEALHLMQRLGDA---WRACLKGYGMWG---NDAAFDVRGAANGD----- 269
Qy      231 AYLTKIASGVMEVSGDGKSGRHGNRAPMEIAVDVAGDPOALEMREFFGSGWG--RR 288
Db      270 -YVAK--WGAEEELTSSSKSGKRGRTPROLL--QAQDD---GLMIEFNATSGKRRR 320
Qy      289 AIAWSRGRLRARGLAGELTD---AQIVEQESAPVWVAIIPASMMMR 334
Db      321 QLVMSQGLKEECGL-VELDDDEAMAEVDAEAGSGPEIVAEMDNBSGKQVR 369
```

```
RESULT 15
US-10-261-942-9
; Sequence 9, Application US/10261942
; Publication No. US20030087440A1
; GENERAL INFORMATION:
; APPLICANT: D'Elia, John
; APPLICANT: Stoddard, Steven F.
; TITLE OF INVENTION: KETOGLONIGENIUM ENDOGENOUS PLASMIDS
; FILE REFERENCE: 1533.1260002
; CURRENT APPLICATION NUMBER: US/10/261,942
; PRIOR FILING DATE: 2002-10-02
; PRIOR APPLICATION NUMBER: US 09/826,191
; PRIOR FILING DATE: 2001-04-05
; PRIOR APPLICATION NUMBER: US 60/194,627
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 9
; LENGTH: 466
; TYPE: PRT
; ORGANISM: Ketogulonigenium
US-10-261-942-9
```

```
Query Match      14.1%; Score 282; DB 15; Length 466;
Best Local Similarity 30.0%; Pred. No. 7.2e-20;
Matches 87; Conservative 40; Mismatches 119; Indels 44; Gaps 15;

Qy      60 NBPKSGFGGLRSCGKGMICPCGAGKVGARADEISQVVAHQLTGGSV-AMVTMTMRHTA 118
Db      109 DGGRRAFFSGLAQCRRNWGCVCASARIADIRSENMHLLAMARDNGFVPLITLTAQHKA 168
Qy      119 GQRLHDLMTGLSAAMKAAATNGRRMTEREMYGCD---GYRAVEITHG-KNGWHVHYHA 173
Db      169 GDSLFDDLQNMKKAQRLRORREWR-----DLPEVGSVTSTETHSYANGMHPHFE 220
```


GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 8, 2003, 22:29:38 ; Search time 34 Seconds
(without alignments)
1071.998 Million cell updates/sec

Title: US-10-007-527A-2

Perfect score: 2005

Sequence: 1 MTSVSAHLSGKRRPVLVS.....HLHYRLPADYRPPIISVRK 379

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283308 segs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR 76:*

1: pirl:*

2: pirl:*

3: pirl:*

4: pirl:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	321.5	16.0	256	2 S34850	hypothetical prote
2	316	15.8	456	2 A31844	rep protein - Stre
3	230.5	11.5	437	2 S04020	hypothetical prote
4	169	8.4	240	2 T40335	hypothetical prote
5	127	6.3	314	2 B35390	replication protei
6	123.5	6.2	315	2 S25721	replication protei
7	122.5	6.1	315	2 T44800	replication protei
8	118.5	5.9	315	2 JN0856	rep protein - Str
9	115	5.7	336	2 A40651	replication protei
10	114	5.7	314	2 T46831	plasmid replicatio
11	111.5	5.6	333	2 A49783	replication initia
12	105	5.2	381	2 A83561	probable type II s
13	104.5	5.2	340	2 T140561	replication initia
14	104	5.2	420	2 AG2059	heme biosynthesis
15	102.5	5.1	420	2 AC3284	D-amino-acid dehyd
16	101.5	5.1	339	2 S01098	rep protein - Bact
17	100.5	5.0	614	2 G75393	hypothetical prote
18	99.5	5.0	614	2 D87410	fold protein [impo
19	98.5	4.9	532	2 H83493	hypothetical prote
20	98.5	4.9	353	2 H83641	probable hemagglut
21	98	4.9	442	2 D87594	genetic exchange p
22	95.5	4.8	606	2 D83484	probable acyl-CoA
23	94.5	4.7	1198	2 T28678	polyketide synthas
24	94.5	4.7	1763	2 T17465	ribose polymerase
25	94	4.7	317	2 A43621	probable replicati
26	94	4.7	382	2 A82861	L-lactate dehydrog
27	94	4.7	382	2 D97338	chaperonin groEL
28	94	4.7	544	1 BVCGL	hypothetical prote
29	94	4.7	1245	2 H87340	hypothetical prote

30	93	4.6	317	2 S25294	replication protei
31	93	4.6	748	2 T37097	probable secreted
32	93	4.6	796	2 S57844	lethal (3) malignant
33	92.5	4.6	354	2 B75355	hypothetical prote
34	92.5	4.6	600	2 B72027	phosphoenolpyruvat
35	92.5	4.6	600	2 A86597	phosphoenolpyruvat
36	92.5	4.6	1026	2 G87346	hypothetical prote
37	92	4.6	799	2 T48889	hypothetical prote
38	91.5	4.6	302	2 H96792	serine/threonine p
39	91.5	4.6	332	2 S51694	unknown protein F1
40	91.5	4.6	350	2 B82281	replication protei
41	91	4.5	296	1 CPNB7	ferric vibriobacti
42	91	4.5	956	2 B83200	peroxidase (EC 1.1
43	90.5	4.5	320	2 D96750	probable phosphotr
44	90.5	4.5	239	2 A83009	unknown protein F2
45	90.5	4.5	2399	2 P98274	non-ribosomal pept
					hypothetical prote

ALIGNMENTS

RESULT 1

S34850
hypothetical protein - Streptomyces lavendulae plasmid pSLG33

C/Species: Streptomyces lavendulae

C/Date: 19-Mar-1997 #sequence_revision 25-Apr-1997 #text_change 07-May-1999

C/Accession: S34850

R/Felsberg, J.; Petricek, M.; Tichy, P.

Nucleic Acids Res. 21, 3582, 1993

A/Title: Nucleotide sequence of the mini-plasmid pSLG33 from Streptomyces lavendulae

A/Reference number: S34850; MUID:93348001; PMID:8346038

A/Accession: S34850

A/Status: preliminary; translation not shown

A/Molecule type: DNA

A/Residues: 1-256 <FEEL>

A/Cross-references: EMBL:X69872

A/Genetics:

A/Genome: plasmid pSLG33

Query Match 16.0%; Score 321.5; DB 2; Length 256;

Best Local Similarity 33.2%; Pred. No. 5.3e-19;

Matches 85; Conservative 45; Mismatches 105; Indels 21; Gaps 11;

QY	109	MVTMTTRHTAQRDLMTGLSAANKAATNGRRM-----TERENYGDGYRAVEI	160
DB	1	MVTTLARHKHHELEPLDAVANGWRKLLSGRAMGDPKRGVLAGRDLGVNGIRSLIEV	60
QY	161	THG-KNGHNVHALLMFSGDVSENIIESPDDAMPDR-WTSKTVSLGFAAPLRNSGGADV	218
DB	61	TYGTNGMHPHLVHLINLETTEEL--AYAMHRMDKTRAWLKKAGPEPSKEH--GITW	116
QY	219	RKIGEADQVLAAYVTIKIAGS-VGMEVSGSGDKSGRGNRAPWE-IAVDVAGDQQLDEL	276
DB	117	SKV--TTPEEAGEYIAKQBGKVGNEIRAGDKAKRLGTLAPFEULEFRQGDVAVYVI	174
QY	277	WREFFSGMGRRAIWSRGLPAR-AGLGAELTDQIVQESAPYVAIIPARSMWMTIRT	335
DB	175	WQEVKGTFRRAITWSRGLRAELLDGDEBELDDELAELIGEGFTW-ALLPESIRAIRR	233
QY	336	CAPYVFGEILGIVENG 351	
DB	234	-VPGIQRILDTAENG 248	

RESULT 2

rep protein - Streptomyces lividans plasmid pIJ101

C/Species: Streptomyces lividans

C/Date: 28-Feb-1990 #sequence_revision 28-Feb-1990 #text_change 23-Mar-1993

C/Accession: A31844

R/Kendall, K.J.; Cohen, S.N.

J. Bacteriol. 170, 4634-4651, 1988

A/Title: Complete nucleotide sequence of the Streptomyces lividans plasmid pIJ101 and

A:Reference number: A91888; MUID:89008081; PMID:3170481
A:Accession: A31844
A:Molecule type: DNA
A:Residues: 1-456 <K>
C:Genetics:
A:Genome: plasmid

Query Match 15.8%; Score 316; DB 2; Length 456;
Best Local Similarity 30.5%; Pred. No. 2.9e-18;
Matches 100; Conservative 40; Mismatches 112; Indels 76; Gaps 15;

QY 49 PISGV-----NGVTIVNGPKSGGFGILRSCGKMGITPCCGAKGVGAHRADEISQVVAHQI 102
DB 3 PASGVTAQTAAGTSVVL-----GLMRCGRIMLCVCAATIRHKRAEETITAAVWMI 54
QY 103 GTGSVA-MVTMTMTTAQORLHDLMTGSAANKATNRR-----WTE----- 145
DB 55 KRGGTAYLVTFARIGHTRLADLMDALQGTREKTPSPRRPGAVQRLITGGTWAGRRAKD 114
QY 146 -----REMYGCDGVRAVEITTHGK-NGMHVHALIMFSGDV-----SENILRSF 189
DB 115 GHRADREGIDRIGVGMIRATEVTVGQINGMHHIALIVVGRTEGERSAKQIVATF 174
QY 190 --SDMFRWTSKVLVSLGPAPLR-----NSGIDVRKIGESAD-QVLAAYLTIA 237
DB 175 EPTGAALDEWQGHMRSV-WTAALRKVNPAFTPDHGHGFKLETERDANDIAEYIAKTQ 233
QY 238 SG-VGMEVSGDGSGRHGNAPMEI--AYDAVGS--DPQA-----LELMREFE 281
DB 234 DKKAPALBLARADLTATGCVNAPPELLGRIDLTGTGTTEDDAAGVSLNNLSRWHEYE 293
QY 282 FSGMGRRAIAMSRLGAPAPAGLGAELTDA 309
DB 294 RATRGRRAIEMTRYIRQMLGIDGDTTEA 321

RESULT 3

S04020
Hypothetical protein - Streptomyces sp. plasmid PSB24.2

C:Species: Streptomyces sp.
C:Date: 07-Jun-1990 #sequence_revision 07-Jun-1990 #text_change 28-Oct-1994
C:Accession: S04020
R:Boletín, A.P.; Sorokin, A.V.; Aleksandrov, N.N.; Danilenko, V.N.; Kozlov, Y.I.
Dokl. Biochem. 283, 260-263, 1985
A:Title: Nucleotide sequence of DNA of the actinomycete plasmid PSB24.2.
A:Reference number: S04020
A:Accession: S04020
A:Molecule type: DNA
A:Residues: 1-437 <BO>
A:Cross-references: EMBL:X03756
A:Note: the authors translated the codon TGG for residue 59 as Arg, CGG for residue 187
A:Note: the source is designated as Streptomyces cyanogenus
A:Note: this paper is a translation of the Russian paper published in Dokl. Akad. Nauk S
C:Genetics:
A:Genome: plasmid

Query Match 11.5%; Score 230.5; DB 2; Length 437;
Best Local Similarity 24.6%; Pred. No. 3.1e-11;
Matches 102; Conservative 43; Mismatches 140; Indels 129; Gaps 19;

QY 55 GVTIVNGPKSGGFG-GLRSCGKMGITPCCGAKGVGAHRADEISQVVAHQI 112
DB 6 GVIFARTAGAAVALGILKCGRIWLCFVCSGQDPAPRSEITEAVVSWLQGGMAYLVTF 65
QY 113 TMRHTAGORLHDLMTL-----SAAKRAATNGRRWRTRREMY 149
DB 66 TARHTAADRLSDMLDLOGTADATGTGRRPGAVQRLITGGAW--AGDKRRKSNQEGIR 123
QY 150 GCDGY--VRAVEITTHGK-NGMHVHALIMFSGDV-----SENILRSF--DAVF-- 194
DB 124 GATIGYGMIRATEVTVGEGAGMHPHIALIVVGRTEGERSAKQIVATF 183
QY 195 DRWTSKVLVSLGPAPLRNSGL--DVRKIGESADQVLAAYLTIASGVG----- 241

DB 184 DRMRRS-----GPAITLASTPGFRPTTGARSPGATGATATGSSSSVRSGRQPPGVHR 239
QY 242 -----MEVSGDGSGRHGN-----RAPWEIADVAGDPO-----ALELW 277
DB 240 QDAGROEPFGAGATRRPGGRPPGQHDVLRITPSRIG-DLMGVREBEAAGSLAWGLDRW 298
QY 278 REFEFGMG-----RRALMSRGL--RAPAGL 302
DB 299 AEYETAVSAGGHRVDPAPAPGPDRRHRGRHRRGPPVPPRRRRFRDGVQIMDRAMWG 358
QY 303 GAELTDQIYEQESAEVWVAIT-----PARSMMTRTCAPY-----VFEGILG 346
DB 359 VGRSLDAVEAVEVGRISMDALGELVQASQSAFLRVLPQETELLYELLG 412

RESULT 4

140335
Hypothetical protein 1 - Corynebacterium glutamicum

C:Species: Corynebacterium glutamicum
C:Date: 12-Aug-1996 #sequence_revision 12-Aug-1996 #text_change 22-Oct-1999
C:Accession: 140335
R:Yamauchi, R.; Terabe, M.; Miwa, K.; Tsuchiya, M.; Takagi, H.; Morinaga, Y.; Nakamori
Agric. Biol. Chem. 50, 2771-2778, 1986
A:Title: Determination of the complete nucleotide sequence of Brevibacterium lactofermentum
A:Reference number: 140334
A:Accession: 140335
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-240 <RS>
A:Cross-references: GB:D00038; NID:g216381; PIDN:BAA0026.1; PID:g216383

Query Match 8.4%; Score 169; DB 2; Length 240;
Best Local Similarity 32.0%; Pred. No. 1.8e-06;
Matches 54; Conservative 28; Mismatches 53; Indels 34; Gaps 10;

QY 193 MFRWTSKVLVSLGPAPLRNSG-GLD-VRKIGESADQVLAAYLTIASGVMEVSGGK 250
DB 1 MFSRWSAGVVAGMDAPLRHGVKLDQVSTWGDAAK-MATYIAK---GMSQELTGSATK 56
QY 251 SGRHGNRAPMEIADVAGDPOALE-----LMREFEFGSMGRALMSRGLRARGL 302
DB 57 TASKSGTYTFQW-LDMLADQSDAGEDMDAVLVAKRREIVSGSKURS-SMSRG--AKGAL 112
QY 303 GAELTDAQI---VEQE-----ESAPVVAIIPARSMMTRT 335
DB 113 GIDYIDAVRREMEBELYKLAGBAPEVSTTRVAVLVKDDWKLIGS 161

RESULT 5

B35390
Replication protein - Lactobacillus hilgardii

C:Species: Lactobacillus hilgardii
C:Date: 31-Aug-1990 #sequence_revision 31-Aug-1990 #text_change 24-Sep-1999
C:Accession: B35390
R:Johnson, K.; Soetaert, P.; Michiels, F.; Joos, H.; Mahillon, J.
J. Bacteriol. 172, 3089-3099, 1990
A:Title: Lactobacillus hilgardii plasmid pLAB100 consists of two functional cassettes
A:Reference number: A35390; MUID:90264300; PMID:2188951
A:Accession: B35390
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-314 <JOS>
A:Cross-references: GB:M55222; NID:g149352; PIDN:AAA98163.1; PID:g149354; GB:M37906
C:Superfamily: replication protein

Query Match 6.3%; Score 127; DB 2; Length 314;
Best Local Similarity 22.0%; Pred. No. 0.0071;
Matches 55; Conservative 50; Mismatches 107; Indels 38; Gaps 10;

QY 75 KGMICPPCAGKAGARADISQV--AH-QLGTGSVAWMTMTTAAQORLHDLMTGSA 131
DB 69 KSRICPLCNWRSMGQSNQMLQVLDAAKQRTGTFPLITTAENASGENLKQVRYKMR 128

[illegible]

RESULT 13

I04561
replication initiation protein rep60 [validated] - Bacillus subtilis plasmids

N|Comments: DNA ligase (EC 6.5.1.-); DNA lyase (EC 4.2.99.-)

C|Species: Bacillus subtilis

C|Date: 12-Aug-1996 #sequence revision 12-Aug-1996 #text_change 20-Apr-2000

C|Accession: I04561, A32059, A43997, I40549

R|Melzer, M.J.; Venema, G.; Bion, S.

Nucleic Acids Res. 23, 612-619, 1995

A>Title: Characterization of single strand origins of cryptic rolling-circle plasmids from

A|Reference number: I40549; MUID:95206941; PMID:7899081

A|Accession: I40561

A|Status: preliminary; translated from GB/EMBL/DDBJ

A|Molecule type: DNA

A|Residues: 1-340 <MEI>

A|Cross-references: EMBL:U03380; NID:g1049123; PID:AAC44418,1; PID:g1049124

A|Experimental source: plasmid pTA1060

R|Devine, K.M.; Hogan, S.T.; Higgins, D.G.; McConnell, D.J.

J. Bacteriol. 171, 1166-1172, 1989

A>Title: Replication and segregational stability of Bacillus plasmid pBA1.

A|Reference number: A32059; MUID:89123116; PMID:2492507

A|Accession: A32059

A|Molecule type: DNA

A|Residues: 34-124, 'V', 126-138, 'E', 140-230, 'L', 232-237, 'E', 239-340 <DEV>

A|Cross-references: GB:M24251; NID:g142426

A|Experimental source: plasmid pBA1

A>Note: The authors' translation is shown for residues 313-333 and not the sequence shown

R|Darabi, A.; Porough, R.; Bhatnagar, G.; Watabe, M.; Goodarzi, G.; Gross, S.C.; Watabe,

Plasmid 22, 281-286, 1989

A>Title: Identification and nucleotide sequence of the minimal replicon of the low-copy

A|Reference number: A43997; MUID:90223229; PMID:2561213

A|Accession: A43997

A|Molecule type: DNA

A|Residues: 34-124, 'V', 126-138, 'E', 140-230, 'L', 232-237, 'E', 239-301, 'R', <DAR>

A|Cross-references: GB:M24386; NID:g144127; PID:AAA23002.1; PID:g144129

A|Experimental source: plasmid pBS2

A>Note: The authors translated the codon ACG for residue 124 as Tyr, and GAG for residue

C|Genetics:

A|Gene: rep60

A|Genome: plasmid

A|Start codon: TTG

C|Superfamily: replication protein

C|Keywords: carbon-oxygen lyase; ligase; plasmid replication

F|253/Active site: Tyr (covalent DNA-binding) #status predicted

Query Match 5.2%; Score 104.5; DB 2; Length 340;
Best Local Similarity 24.2%; Pred. No. 0.56;
Matches 38; Conservative 35; Mismatches 57; Indels 27; Gaps 9;

OY 78 ICPCCAKVG--AHRADISQVAHQLGTGSVAMVTWTKRHAYGQRLLHDIMTGLSAAMK 134
:::|||||||
87 LCPMCAMRRSLIKIVHNKLVEEANROYGCWMI-FILITINVGGRLEKPOLISAMBEGR 145
:::|||||

```

QY 135 AAINRGRRTERERMGCOCDOYAVAVIT--HGNGGHHVHALL--MFGSDVSENTEBESFD 139
Db 146 KLFQYKVKVT-----SVSGFRALLETIKKHEDDYHMFHVLIVKRNKYNFNTIKQ--- 197
QY 192 AMFDKMTS---KVLISGFAPLDRNSGDLVDRKIGSEA 225
Db 198 --AEWTSLSMKKAMRLDY--TPI-----VDIRVKKKA 225

```

[illegible]

A:Molecule type: DNA
 A:Residues: 1-420 <KUR>
 A:Cross-references: GB:AE008917; PIDN:AAL51438.1; PID:g17982146; GSPDB:GN00190
 A:Experimental source: strain 16M
 C:Genetics:
 A:Gene: BME10256
 A:Map position: 1
 C:Keywords: oxidoreductase

Query Match 5.1%; Score 102.5; DB 2; Length 420;
 Best Local Similarity 19.9%; Pred. No. 1;
 Matches 63; Conservative 49; Mismatches 111; Indels 93; Gaps 12;

Qy	95	SCVVAHQLOTGTSVAVMTMTMHTAGQRLHDLWTGLSAWKAATNGRRMR--TEREMYGCD	152
Db	152	AEIAASQPGMDVREKAGIAVEHVRGRRLAEIQGLNPRFVAGTFVPGWKNVSDPKLF---	208
Qy	153	GYRAVEITHGKGMHVHVALLMFSGDVSENILESFSDDAMPDRMTSKLVSLGFAPAPLRN	212
Db	209	-----GKAIW-----SYAESLGARFLSGKV---ASAKRQ	234
Qy	213	SGGLDVKRTIGGEADQVLAAYLTKIASGVGMEYSGDGKSGRHNRAPMELAVDAVGDPQ	272
Db	235	NGGVRLRLENGV--EINATHLVLMAGAMSRDLAKPG-----DIVPLDTE	277
Qy	273	-----ALELWREFFEGSGMGRRAIAMSRLPARAGL---GABL-----TDAQIVE	313
Db	278	RGYNTTLPYGSFDPVKRQLTFPGHFVITPMETGLRVGAVEFGGLDLPENFARSEAMLKX	337
Qy	314	QESAPVWVAITPARSMWMITCAP-----YVFGG-IIGLVEAGATMEN	356
Db	338	ASKFLPGL-KVEGGRQMGMGRPSMPDLPVIGRASAGNYYVYGFHGH/LGLOSAAATATL	396
Qy	357	LRDHLHYRLPADVRP	372
Db	397	IRDLITGSEPAIDIEP	412

Search completed: November 8, 2003, 23:12:53
 Job time : 36 secs

Db 175 EPTGALDDEMGHMSV-WTAAIKRNPAFTPDHRCGVDFPKLETRDNDLAEYIAKTQ 233
QY 238 SG--VQMEVSGDGSGRGNRAPWEI---AVDAVG3---DPOA-----LELMREFE 281
Db 234 DGRAPALBELRADLTKATGAGNAPFELLRIGDLTCGMTEDDAAGVGLJEMNLNRWHEZE 293
QY 282 FCGMGRRAIAMSFGLEPARAGLGAELTDA 309
Db 294 RAIRGRRAIEMTRYLRQMLGLDGDTEA 321

RESULT 2

NEOR_STRCY
ID NEOR_STRCY STANDARD; PRT; 437 AA.
AC P14501;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JUN-1990 (Rel. 13, Last sequence update)
DT 01-NOV-1990 (Rel. 16, Last annotation update)
DE Neomycin resistance protein.
OS Streptomyces cyanogenus.
OC Plasmid pSB24.2.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomycinae; Streptomycetaceae; Streptomyces.
OX NCBI_Taxid=80860;
RN [1]
RP SEQUENCE FROM N.A.
RA Bolotin A.P., Sorokin A.V., Aleksandrov N.N., Danilenko V.N.,
RA Kozlov Y.I.;
RT "Nucleotide sequence of DNA of the actinomycete plasmid pSB24.2";
RL Dokl. Biochem. 283:1014-1017(1985).

CC [2]
CC SEQUENCE FROM N.A.
CC Bolotin A.P., Sorokin A.V., Aleksandrov N.N., Danilenko V.N.,
CC Kozlov Y.I.;
CC "Nucleotide sequence of DNA of the actinomycete plasmid pSB24.2";
CC Dokl. Biochem. 283:260-263(1985).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL; X03756; CAA27389.1; -
CC EMBL; M32513; AA98338.1; -
CC InterPro; IPR000989; Rep.
DR Pfam; PF01446; Rep; 1.
KW Antibiotic resistance; Plasmid.
CC CONFLICT 111 111 G -> R (IN REF. 2).
FT CONFLICT 220 220 A -> P (IN REF. 2).
FT CONFLICT 356 356 K -> N (IN REF. 2).
FT CONFLICT 412 412 R -> G (IN REF. 2).
SQ SEQUENCE 437 AA; 47409 MW; 2AF299AD04EB01B CRC64;

Query Match 11.3%; Score 227; DB 1; Length 437;
Best Local Similarity 24.9%; Pred. No. 2, 4e-11;
Matches 94; Conservative 35; Mismatches 126; Indels 122; Gaps 17;

QY 55 GTTVNGPKSGFG-GLRSCGKMWICPCAGKVGARADEISQVVAHOLGTGSVA-MVTM 112
Db 6 GYIFARTAGAANAVALGLTKCGRIWLCPCVSGQDPAPRSEITBAVVSMLOCGGMAYLVTF 65
QY 113 TMRHTAAGRLHDLMTGL-----SAAMKATNGRRNRTEREY 149
Db 66 TARHTAADLSDLMALQGTADAETGTRRPGAYORLITGAAM--AGDKRRKSNOEGIR 123
QY 150 GGDGY---VRAVEITGKN-GWVHVHALIMFSGV-----SENILESPS---DAWF--- 194
Db 124 GRIGTIGMIRATEVIVGEGAGMHPHIALIVGRTGEGRGDKRIITGITPSBDALTEWE 183

QY 195 DRWTSKLVSLGRAPLRNSGL-----DYRKIGERADQVLAAYLTIASGV----- 241
Db 184 DRWRS-----GPATLARSTPGFRPTGARSPPATGATATASSTSSSVSGORPGRVHR 239
QY 242 -----MEVSGDGSGRGN-----RAWMEIAYDAVGGDPQ-----ALELM 277
Db 240 QDAGQEPGPGAGTRRPGQPHDVLKTPSRIG-DLMGGVPEEAAHGSLAWGLDRW 298
QY 278 REFPGNSG-----RAI-----AMSRGLRAR 299
Db 299 AEYETAAGAGHRVPLPAPAGPDRRRHRRGRHGRFPVDRRRRRFRGQVQWDAMKGL 358
QY 300 AGLGAELTDAQIVGQEE 316
Db 359 VGRSLDLAVVEVRE 375

RESULT 3

REP_LACHI
ID REP_LACHI STANDARD; PRT; 314 AA.
AC P35857;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Rep protein (Replication protein).
GN REP.
OS Lactobacillus hilgardii.
OG Plasmid pLAB1000.
OC Bacteria; Firmicutes; Lactobacillales; Lactobacillaceae;
OC Lactobacillus.
OX NCBI_Taxid=1588;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=90264300; PubMed=2188951;
RA Jossan K., Soetaert P., Michiels F., Joos H., Mahillon J.;
RT "Lactobacillus hilgardii plasmid pLAB1000 consists of two functional
RT cassettes commonly found in other Gram-positive organisms.";
RL J. Bacteriol. 172:3089-3099(1990).
CC -1- FUNCTION: PRODUCES A SINGLE-STRAND NICK IN A SPECIFIC SITE OF THE
CC PLASMID, AND THIS NICK RESULTS IN SINGLE-STRAND REPLICATION BY
CC ROLLING CIRCLE MECHANISM.
CC -1- SIMILARITY: TO SEVERAL PROTEINS INVOLVED IN GRAM-POSITIVE BACTERIA
CC PLASMIDS REPLICATION.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL; M55222; AAA98163.1; -
CC EMBL; A14660; CAA01184.1; -
CC PIR; B35390; B35390.
DR InterPro; IPR000989; Rep.
DR Pfam; PF01446; Rep; 1.
KW DNA replication; Plasmid.
FT BINDING 229
SQ SEQUENCE 314 AA; 37044 MW; 95144612185988B CRC64;

Query Match 6.3%; Score 127; DB 1; Length 314;
Best Local Similarity 22.0%; Pred. No. 0.003;
Matches 55; Conservative 50; Mismatches 107; Indels 38; Gaps 10;

QY 75 KMWTPCCAGKVGARADEISQV--AH-QLTGSVAAVMTMRHTAAGRLHDLMTGISA 131
Db 69 KSRLLPPLNWRSMQSQNLQVLDGAHKORTGFFLTLTAENASGENTKQEVKMGK 128
QY 132 AKKATNGRRRTEEMTGCGGYVAAVITGKNG-WVHVHALIMFGSDVSENILESPS 190
Db 129 A-----ISKLPYKPKAPKMLGIVASTETITNKGTHQHWAVLTFVKPTFKDSANYTN 183

```

DT      01-JUN-1994 (Rel. 29, Last annotation update)
DE      Rep protein (Replication protein).
GN      REP.
OS      Bacillus amyloliquefaciens.
OG      Plasmid pFTB14.
OC      Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX      NCBI_TaxID=1390; [1]
RN      RP
RP      SEQUENCE FROM N.A.
RX      MEDLINE=86121709; PubMed=3481020;
RA      Murai M., Miyashita H., Araki H., Seki T., Oshima Y.;
RT      "Molecular structure of the replication origin of a Bacillus
RL      amyloliquefaciens plasmid pFTB14."
RM      Mol. Gen. Genet. 210:92-100(1987).
CC      -1- SIMILARITY: TO SEVERAL PROTEINS INVOLVED IN GRAM-POSITIVE BACTERIA
CC      PLASMIDS REPLICATION.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL; X06242; CAA29586.1; -.
DR      PIR; S01098; S01098.
DR      InterPro; IPR000989; Rep.
DR      Pfam; PF01446; Rep; 1.
KW      DNA replication; Plasmid.
FT      BINDING 251
SQ      SEQUENCE 339 AA; 39641 MW; F24056B961E28DE7 CRC64;
Query Match          5.1%; Score 101.5; DB 1; Length 339;
Best Local Similarity 23.8%; Pred. No. 0.41;
Matches 39; Conservative 37; Mismatches 57; Indels 31; Gaps 10;
OY      78 ICPCGAGVVG---AHRADETISQVVAHQUGTSGSVAMVTMTMRHTTGQRHDLMTGLSAAWK 134
DB      86 LCPWCAMFRSLKTIHYNHKLITKANRQGGCWI-PLTLTVANXGKERLKPOISEMMEGFR 144
OY      135 AATNGRRWRTERENYGCDGYRAVEIT--HGKNQHVVHVALMLPSCD-VSENILSPSD 191
DB      145 KLFGYKKVKYT-----SVLGFPRALEITNQHEDTYHPHFVLLPVKRYPKGNKYIKQ--- 196
OY      192 AMFPRWTS---KLVS LGFAPLRNSGGLDVARKIGS---EADV 228
DB      197 --AWTSTLMKRANKLDY-FPI-----VDIRRVGRVKAIDAQI 231
RESULT 6
REP_LACPU
ID      REP_LACP1 STANDARD; PRF, 317 AA.
AC      PI6953;
DT      01-AUG-1990 (Rel. 15, Created)
DT      01-AUG-1990 (Rel. 15, Last sequence update)
DT      16-OCT-2001 (Rel. 40, Last annotation update)
DE      Rep protein (Replication protein).
GN      REP.
OS      Lactobacillus plantarum.
OG      Plasmid pUP1.
OC      Bacteria; Firmicutes; Lactobacillales; Lactobacillaceae;
OC      Lactobacillus.
OX      NCBI_TaxID=1590; [1]
RN      RN
RP      SEQUENCE FROM N.A.
RX      STRAIN=ATCC 8014 / CCM 1904;
RX      MEDLINE=90223315; PubMed=2517345;
RA      Bouia A., Bringel F., Frey L., Kammerer B., Belarbi A.,
RA      Guyonvarch A., Hubert J.-C.;
RT      "Structural organization of pUP1, a cryptic plasmid from
RT      Lactobacillus plantarum CCM 1904." ;
RT      Plasmid 22:185-192(1989).

```

```
CC -!- FUNCTION: PRODUCES A SINGLE-STRAND NICK IN A SPECIFIC SITE OF THE
CC PLASMID, AND THIS NICK RESULTS IN SINGLE-STRAND REPLICATION BY
CC ROLLING CIRCLE MECHANISM.
CC -!- SIMILARITY: TO SEVERAL PROTEINS INVOLVED IN GRAM-POSITIVE BACTERIA
CC PLASMIDS REPLICATION.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M31223; AAA8164.1; -.
CC DR InterPro; IPR000989; Rep.
CC DR Pfam; PF01446; Rep; 1.
CC KM DNA replication; Plasmid.
CC FT BINDING 229 229 DNA (BY SIMILARITY).
CC SQ SEQUENCE 317 AA; 37009 MW; AD9A9B594FF73DC CRC64;

Query Match 4.9%; Score 99; DB 1; Length 317;
Best Local Similarity 19.7%; Pred. No. 0.62;
Matches 58; Conservative 50; Mismatches 115; Indels 72; Gaps 12;

QY 70 LRSCG-----KGM-----ICPCAGKVGARADEISQVVAHQIG--TGS 106
DB 42 VRGCGEVLNFRKIGETHLKLYQTFCFKRLCPICNMRSMKNSQLKQIIAEVAREPKR 101
QY 107 VAMVMTWR--HQAQRRLHDLMTGLSAWKAKATNGRRM--TEREYGGCDGYRAVEIH 162
DB 102 FLFLTLTVKNAHSA---ELKVSLPALTKAFNKLTRYKRVTKNL---GYLRSTETIV 153
QY 163 GK--NGMHVHALIMFSGDVSENILSFSDAMFDEMTSKLVSLGFAPLRNSGLDVRK 220
DB 154 NEQDGSYNQHLVLFVKSSYKNSNYLAQAEKAKLQKALKVDEPVAHYQAYKAKR 213
QY 221 IGGEADQVLAAYLTIKIASGVMEVSGSGSRHGNRAPEIAVDAVGDPQALMEREF 280
DB 214 KETDSIQASAEETAK-----YEKSDADYMT-----ADDERNLVYIKNL 251
QY 281 EFGSMGRALIMSRGIRARAIGLAEITDAQ-----IVGESEAPVWVA 323
DB 252 EYALAGTRQISYG-GILKQIKODLKEDVENGDVHVHVEDYTKEQMEAAEEVVA 305

RESULT 7
PURL_PYRAE STANDARD; PRT; 697 AA.
AC Q8Z207,
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Phosphoribosylformylglycinamide synthase II (EC 6.3.5.3) (FGAM
DB synthase II).
GN PURL OR PA80225.
OS Pyrobaculum aerophilum.
OC Archaea; Crenarchaeota; Thermoprotei; Thermoproteales;
OC Thermoproteaceae; Pyrobaculum.
OX NCBI_TaxID=13773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=IM2 / ATCC 51768 / DSM 7523;
RX MEDLINE=2164397; PubMed=11792869;
RA Pitt-Gibson S.T., Ladner H., Kim U.-J., Stetter K.O., Simon M.I.,
RA Miller J.H.,
RT "Genome sequence of the hyperthermophilic crenarchaeon Pyrobaculum
RT aerophilum."
RT Proc. Natl. Acad. Sci. U.S.A. 99:984-989(2002).
CC -!- CATALYTIC ACTIVITY: ATP + 5'-phosphoribosylformylglycinamide + L-
CC glutamine + H(2)O = ADP + phosphate + 5'-
CC phosphoribosylformylglycinamide + L-glutamate.
CC -!- PATHWAY: De novo purine biosynthesis; fourth step.
```

```
CC -!- SUBUNIT: Heterodimer of two subunits, purQ and purl.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -!- SIMILARITY: BELONGS TO THE FGMS FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AB09757; AAL2642.1; -.
CC DR HAMAP; MF_00420; -; 1.
CC DR InterPro; IPR00728; AIRS_related.
CC DR Pfam; PF00586; AIRS; 2.
CC DR Pfam; PF02769; AIRS_C; 2.
CC KM Purine biosynthesis; ligase; ATP-binding; Complete proteome.
CC FT NP_BIND 91 102 ATP (POTENTIAL).
CC SQ SEQUENCE 697 AA; 74373 MW; 94219332A19B8558 CRC64;

Query Match 4.7%; Score 95; DB 1; Length 697;
Best Local Similarity 20.8%; Pred. No. 3.2;
Matches 52; Conservative 34; Mismatches 94; Indels 70; Gaps 10;

QY 47 GRPIGVNGVYTVNGPKSGF-----GLRSCGKGIICPCAGKVGARADEISQVVA 99
DB 465 GRFLAAYDSINVGSPSPSAWQFEVAEGRBAEALGVIVGKVSIVNEYKGPVAP 524
QY 100 HQLGTSVAMVTMKRHTAGRLHDLMTGLSAWKAKATNGRRMTEREYGGCDGYRAVE 159
DB 525 -----TVAVVVLGV-----VEDVSKNRAVMEGQGVYVGTKDEVGGSSEYLH--- 568
QY 160 ITHGKNHVAHVHALIMFSGDVSENILSFSDAMFDEMTSKLVSLGFAPLRNSGLDVR 219
DB 569 -----RVHGIVAQPPSIDYSVEKEIAAIVKRAAGRL-----TAKDV- 606
QY 220 KIGGEADQVLAAYLTIKI--ASGVGNEV-----GSGDGKSGRHGNRAPW 260
DB 607 GIGG-----LAAALAKMAAAGIGADIDICAPSTTARLDYLLPESNGRFLAAGEBG- 660
QY 261 EIAVDAVGSD 270
DB 661 GTRVGAAGSD 670

RESULT 8
ARGD_BIFLO STANDARD; PRT; 431 AA.
AC P59315,
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Acetylornithine aminotransferase (EC 2.6.1.11) (ACOAT).
DB ARGD OR BL1061.
GN Bifidobacterium longum.
OS Bacteria; Actinobacteriae; Actinobacteriales; Bifidobacteriales;
OC Bacteria; Actinobacteriae; Bifidobacterium.
OX NCBI_TaxID=216816;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NCC 2705;
RX MEDLINE=22294977; PubMed=12381787;
RA Schell M.A., Karimantzou M., Snel B., Vilianova D., Berger B.,
RA Pessi G., Zwaalen M.-C., Desiere F., Bork P., Delley M.,
RA Priemore R.D., Arigoni F.,
RT "The genome sequence of Bifidobacterium longum reflects its adaptation
RT to the human gastrointestinal tract."
RT Proc. Natl. Acad. Sci. U.S.A. 99:14422-14427(2002).
CC -!- CATALYTIC ACTIVITY: N(2)-acetyl-L-ornithine + 2-oxoglutarate = N-
CC acetyl-L-glutamate + 5-semialdehyde + L-glutamate.
CC -!- CORCTOR: Pyridoxal phosphate (By similarity).
CC -!- PATHWAY: Arginine biosynthesis; fourth step.
```

```

CC -1- MISCELLANEOUS: May also have succinylidiaminopinelate
CC aminotransferase activity, thus carrying out the fourth step in
CC lysine biosynthesis.
CC -1- SIMILARITY: Belongs to class-III of pyridoxal-phosphate-dependent
CC aminotransferases.
CC -----
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC DR EMBL; A014729; AAN24869.1; -
CC DR HAMAP; MF 01107; -; 1.
CC DR Pfam; PF00202; aminotran_3; 1.
CC DR TIGRfams; TIGR00707; argd; 1.
CC DR PROSITE; PS00600; AA_TRANSFEX_CLASS_3; 1.
CC KW Arginine biosynthesis; Transferase; Aminotransferase;
CC PYridoxal phosphate; Complete proteome.
CC FT BINDING 284 284 PYRIDOXAL PHOSPHATE (BY SIMILARITY).
CC SQ SEQUENCE 431 AA; 45615 MW; 1A9245B103521BCE CRC64;

Query Match 4.7%; Score 94.5; DB 1; Length 431;
Best Local Similarity 22.9%; Pred. No. 2.1;
Matches 50; Conservative 27; Mismatches 72; Indels 69; Gaps 9;

QY 134 KAATNGRRMREEMTGCDGY--VRAVEITHGKNGMHVHVALMFSGDVSENILSPS 190
DB 8 KLTGDSKMLGEGSYQHMANVFGTFLAYMDHGGCAHINVDGNEYLDLAGIAYN----- 61
QY 191 DAMFDRTSKLVSGLPAAPLRNSGLDVRKIGEDQV-----IAAYLTXI 236
DB 62 -----SLGYAHFKM-----VKAVDQAQKVAHISNYFASPEQIELASGLVLT 103
QY 237 ASGVMEVGG--DGSGRGHGNAPMEI-----AYDAVGDP-QALMLNRFEEG 283
DB 104 A---GAPEGSKVYFGNSGARGNAALAKLAKLYRTLPGLPSTIGKPARILMLTHGFHGR 160
QY 284 SMGRRAIAMSRLR-----AAAGLGAELTDA 309
DB 161 TWGALSAWKPKGIRKPYDPLVPNIFFVRAGDKVALHDA 198

RESULT 9
CH60 SYNP7 STANDARD; PRT; 544 AA.
ID CH60_SYNP7 STANDARD; PRT; 544 AA.
AC P22879;
DT 01-AUG-1991 (Rel. 19, Created)
DT 01-AUG-1991 (Rel. 19, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE 60 kDa chaperonin (Protein Cpn60) (groEL protein).
GN GROL OR GROEL OR WOPR OR CPN60.
OS Synecococcus sp. (strain PCC 7942) (Anacystis nidulans R2).
OC Bacteria; Cyanobacteria; Chroococcales; Synecococcus.
OX NCBI_TaxID=1140;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=90368561; PubMed=1975581;
RA Webb R., Reddy K.J., Sherman L.A.;
RT "Regulation and sequence of the Synecococcus sp. strain PCC 7942
RT groEL operon, encoding a cyanobacterial chaperonin."
RL J. Bacteriol. 172:5079-5088(1990).
CC -1- FUNCTION: Prevents misfolding and promotes the refolding and
CC proper assembly of unfolded polypeptides generated under stress
CC conditions (By similarity).
CC -1- SUBUNIT: Oligomer of 14 subunits composed of two stacked rings of
CC 7 subunits (By similarity).
CC -1- STRUCTURAL LOCATION: Cytoplasmic (By similarity).
CC -1- SIMILARITY: Belongs to the chaperonin (HSP60) family.
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration

```

```

CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC DR EMBL; M58751; AAA27314.1; -
CC DR PIR; B36721; BYVCGL.
CC DR HSSP; P06139; IGR1.
CC DR HAMAP; MF 00600; -; 1.
CC DR InterPro; IPR001844; Chaprin Cpn60.
CC DR InterPro; IPR002423; Cpn60/TCP-1.
CC DR Pfam; PF00118; Cpn60_TCP1; 1.
CC DR PRINTS; PR00298; CHAPERONIN60.
CC DR PRINTS; PR00304; TCOMPLEXTCP1.
CC DR PROSITE; PS00296; CHAPERONIN_CPN60; 1.
CC KW Chaperone; ATP-binding.
CC SQ SEQUENCE 544 AA; 58096 MW; 29665F9A57D190FB CRC64;

Query Match 4.7%; Score 94; DB 1; Length 544;
Best Local Similarity 21.5%; Pred. No. 2.9;
Matches 77; Conservative 37; Mismatches 98; Indels 146; Gaps 15;

QY 13 DRPVLYSSDRKGRHRLRPLQOITTSFTNACGP-----ISGVNG 55
DB 215 DEPIILITDKKIGVQDLVPLVEQVAA-----GREPLVIADIEKALATLVNVLNG 268
QY 56 VTIVNPGKSGFGGLRSCGKGMICPCGAGVGAHRADEISQVAHOGT-----GSAVM 109
DB 269 VLVNAAKARAFGFGRRKAMEDIAVLVTGGQ-----TEDAKRKIDITKLDOLEARR 321
QY 110 VTMTFRHTAGRLDLMTGLSAWKAAATNGRRMREEMTGCDGYRAVEITHGKNGMHV 169
DB 322 ITITKQNT-----TIVAEQEAIVKAR-----VQIRQISET----- 354
QY 170 HVHALLMFSGDVSENILSPSDAMPDRTWSKLVSLGAAPLRNSGGLDVKKIGEDQVL 229
DB 355 -----ESSYDK--EKLQRLAKL-----SGVAVVKVGAATTEM 387
QY 230 -----AAVLTIASGVMEVGGSDGSGHGNAPMEIAYDAVGDPQALMLNRFEEG 280
DB 388 KDRRLRLIEDALNAKRAVEGIVPGGTTLA-----HLAP-----QLEEM--- 427
QY 281 EFGSMGRRAIAMSRLRAGL-GAELTDQIVYEQ-----ESAPVVAIIPAR 328
DB 428 -----ATANLSEGLTGAQIVARALTAIRIKRIAEVAGINCAVTSER 468

RESULT 10
MEN_DROME STANDARD; PRT; 796 AA.
ID MEN_DROME STANDARD; PRT; 796 AA.
AC P52302;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Lethal(3)malignant blood neoplasm-1 protein (mbn(83)).
GN L(3)MEN OR L(3)MEN-1.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=Oregon-R;
RA Konrad L., Becker G., Schmidt A., Klockner T.,
RA Kafer-Stillger G., Dreschers S., Egstroom J.-E., Gateff E.;
RT "Cloning, structure, cellular localization, and possible function of
RT the tumor suppressor gene lethal(3)malignant blood neoplasm-1 of
RT Drosophila melanogaster."
RL Dev. Biol. 163:98-111(1994).

```

```

CC - FUNCTION: REQUIRED FOR DIFFERENTIATION OF THE PHAGOCYTIC BLOOD-
CC CELL TYPE, THE PLASMA CYTOTOXIC.
CC - SUBCELLULAR LOCATION: CYTOPLASMIC; MAINLY AROUND THE NUCLEUS.
CC - TISSUE SPECIFICITY: BLOOD CELLS AND OTHER TISSUES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: Z47722; CAAB7654.1; -
DR PIR: S57844; S57844.
DR FlyBase: FBgn0002440; 1(3)mbn.
DR InterPro: IPR000618; Insect cuticle.
DR Pfam: PF00379; Chitin_bind_4; 1.
KM Repeat.
FT DOMAIN 225 314 2 X 37 AA REPEATS.
FT REPEAT 225 261 1-1.
FT REPEAT 278 314 1-2.
FT DOMAIN 471 512 2 X 21 AA REPEATS.
FT REPEAT 471 491 2-1.
FT REPEAT 492 512 2-2.
FT DOMAIN 576 711 7 X APPROXIMATE TANDEM REPEATS,
FT REPEAT 576 598 GLY-SER-RICH.
FT REPEAT 599 618 3-1.
FT REPEAT 619 640 3-2.
FT REPEAT 641 658 3-3.
FT REPEAT 676 694 3-4.
FT REPEAT 695 711 3-5.
FT REPEAT 711 711 3-6.
SQ SEQUENCE 796 AA; 83348 MW; 6B1D593961FEF370 CRC64;

Query Match 4.6%; Score 93; DB 1; Length 796;
Best Local Similarity 23.9%; Pred. No. 5.5;
Matches 80; Conservative 39; Mismatches 108; Indels 108; Gaps 18;

QY 5 SAHLTSG-KDRPVLVSSDKRGIRHELRPLQOITTEPFNACGRISGVNGVTVNGPK 63
DB 522 SAKTLANTKTRPP-----HTFNHQTPILSSAT-----APGISVT-ANTPT 562
QY 64 G-----SGFGIRSGCGKMWICPCAGK--VGNRADEISQVVAHQLOTGSVAVMTWMT 117
DB 563 GNPSPNGGI-AGKAPGNPQAGSGGIIG-APGGRKVSAGISGSAL----- 611
QY 118 AGQRHDLMTGLSAAMKATNGRMRTEREMVCGDGVVAVEITHKXGMYHVALLMF 177
DB 612 -----GGVSGSGKASGNGAIGSGNAIGGATGSKASGFGFSN-----I 651
QY 178 SGDVSENIILSP-SDAMFDEMTSKVLVIGFAFLRNSGGLDVYKTIIGEGADQVLAAYLTKI 236
DB 652 GGGVSGSGKSPGSGESKIGGADGSKASGSGSGSKIGGIGTK----- 695
QY 237 ASGVGMEVSGSGDKS-----GRHKRAWETAVDAVGGDDPPA 273
DB 696 ASGGIGRIGSGRSASATDLYKFIYLDYNGHEBTGNGDK---QSYPAIGED-A 750
QY 274 LELMRFP---EFGSGRRALAWSRGLRARAGLGA 305
DB 751 VQRTIETIANEFQFPH-VSW-RKLDKALPEE 782

```

```

DE carboxykinase) (Phosphoenolpyruvate carboxylase) (PEPCK).
GN PCKG OR CPN0851 OR CP1018.
OS Chlamydia pneumoniae (Chlamydia pneumoniae).
CC Bacteria; Chlamydiae; Chlamydiales; Chlamydiaceae; Chlamydia.
CC NCBI TaxID=83558;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CWL029;
RX MEDLINE=9920606; PubMed=10192388;
RA Kaiman S., Mitchell W., Marathe R., Lammel C., Fan J., Hyman R.W.,
RA Olinger U., Gilmwood J., Davis R.W., Stephens R.S.;
RT "Comparative genomes of Chlamydia pneumoniae and C. trachomatis.";
RL Nat. Genet. 21:385-389(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=AR39;
RX MEDLINE=20150255; PubMed=10684935;
RA Read T.D., Brumham R.C., Shen C., Gill S.R., Heidelberg J.F.,
RA White O., Hickey E.K., Peterson J., Uetzerback T., Berry K., Bass S.,
RA Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,
RA Gwin M., Nelson W., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.,
RA Eisen J., Fraser C.M.;
RT "Genome sequences of Chlamydia trachomatis Mopn and Chlamydia
RT pneumoniae AR39.";
RL Nucleic Acids Res. 28:1397-1406(2000).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=J138;
RX MEDLINE=20330349; PubMed=10871362;
RA Shirai M., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,
RA Shiba T., Ishii K., Hattori M., Kunata S., Nakazawa T.;
RT "Comparison of whole genome sequences of Chlamydia pneumoniae J138
RT from Japan and CWL029 from USA.";
RL Nucleic Acids Res. 28:2311-2314(2000).
CC -! CATALYTIC ACTIVITY: GTP + oxalacetate = GDP + phosphoenolpyruvate
CC + CO(2).
CC -! PATHWAY: Rate-limiting gluconeogenic enzyme.
CC -! SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -! SIMILARITY: BELONGS TO THE PHOSPHOENOLPYRUVATE CARBOXYKINASE [GTP]
CC FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AE01666; AAD18989.1; -
DR EMBL: AE002559; AAF73723.1; -
DR EMBL: AP002548; BAA99059.1; -
DR PIR: A86597; A86597.
DR PIR: E72027; E72027.
DR PHCI-2DPAGE; Q9Z755; -
DR TIGR: CP1018; -
DR HAMAP: MF_00452; -; 1.
DR InterPro: IPR000364; PEP_carboxykin.
DR Pfam: PF00821; PEPCK_1.
DR ProDom: PD004738; PEPCK_N; 1.
DR PROSITE: PS00505; PEPCK_GTP; 1.
KM Gluconeogenesis; lysase; Decarboxylase; GTP-binding; Complete proteome.
FT NP BIND 216 223 GTP (POTENTIAL).
FT ACT SITE 267 267 BY SIMILARITY.
SQ SEQUENCE 600 AA; 66991 MW; 79590DC1A0A9932A CRC64;

Query Match 4.6%; Score 92.5; DB 1; Length 600;
Best Local Similarity 20.0%; Pred. No. 4.4;
Matches 82; Conservative 39; Mismatches 147; Indels 141; Gaps 18;

QY 34 LQOITSTFPNAC---GRPIG-GVNGVTVNGPK-----SGFGGLRS 72
DB 160 LRSIGSGFKLCHSVGKPLSPGADVSPNCPKRSNRIVHFDSDSVMSFGSGYGNAL 219

```

```

QY 73 CGKMTIPCAGKVGARHRADEISQVVAHOLGTG-----SVAMVTMT 113
DB 220 LGR-----KCVALARLASYAKSGOMLAHEHLLIIGITNPEKKYKFSAPSACGKTNLML 275
QY 114 MRRTAQRLHDLMTGLSAAM-KAATNGRRTREYKGCDDYVRAVEITHGNK----- 167
DB 276 MPTLPMKXIECI--GQDIAMIRGRDRYAANPE-YGFPGVAPGTSERTNNALATCRS 332
QY 168 -HVHVHALLMFSGDV-SENILBSFSDAMPDRMTSKVLSG-----FAAPLRNSGG 215
DB 333 NSLFTVVALTADGDVWMEGLTEQPPPLTD-WLGEKWKSGSPAAHPNSRFLAPLRQPS 391
QY 216 LDVRIKIGADQVLAAYLTIKASGVMEVSGDGSGRHNPAPWEIADV--GGDPQA 273
DB 392 LDPEW-----NSPQGVPLDAIIFG----- 411
QY 274 LEHMFEPFGSMKRAIAMSRLGAPAGELTDQIYVQESAPVAVAIIPASMMI 333
DB 412 ---RSEETPLVYELASWEHGVLTIGAGMSSTTAAIVGQLKLRHDPFAMLPFGYNN- 466
QY 334 RTCAFYVFGELGLVE-----AGATW-----ENTR 358
DB 467 ---AYFQHWLSFAENRSLKLPKIFGVWPFKKNNGEFLWPGFSENLR 511

RESULT 12
AFSK STRCO STANDARD; PRT; 799 AA.
ID AFSK STRCO 09F741; 09L002;
DT 01-OCT-1996 (Rel. 34, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DN Serine/threonine protein kinase afsk (EC 2.7.1.37).
GN AFSK OR SC04423 OR SC6F11.21 OR SC06.01.
OS Streptomyces coelicolor.
OC Bacteria; Actinobacteria; Actinomycetales;
OC Streptomycinae; Streptomycetaceae; Streptomycetes.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RX MEDLINE=94341568; PubMed=8063104;
RA Matsumoto A., Hong S.K., Ishizuka H., Horinouchi S., Beppu T.;
RT "Phosphorylation of the afsk protein involved in secondary metabolism
in Streptomyces species by a eukaryotic-type protein kinase.";
RL Gene 146:47-56(1994).
[2]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RX MEDLINE=96186909; PubMed=8635757;
RA Ueda K., Uneyama T., Beppu T., Horinouchi S.;
RT "The aerial mycelium-defective phenotype of Streptomyces griseus
resulting from A-factor deficiency is suppressed by a Ser/Thr kinase
of S. coelicolor A3(2).";
RL Gene 169:91-95(1996).
RN [3]
RP REVISIONS TO 239-240.
RA Matsumoto A., Hong S., Ishizuka H., Horinouchi S., Beppu T.,
RA Uneyama T.;
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2)/M145;
RX MEDLINE=21996410; PubMed=12000953;
RA Bentley S.D., Chater K.F., Cerdano-Tarraga A.-M., Challis G.D.,
RA Thompson N.R., James K.D., Harris D.B., Quail M.A., Kleser H.,
RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
RA Huang C.-H., Kleser T., Larke L., Murphy L., Oliver K., O'Neill S.,
RA Rabinowitsch B., Rajandream M.A., Rutherford K., Rutter S.,
RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
RA Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,

```

```

RA Hopwood D.A.;
RT "Complete genome sequence of the model actinomycete Streptomyces
coelicolor A3(2).";
RL Nature 417:141-147(2002).
CC -1- FUNCTION: INVOLVED IN THE REGULATION OF SECONDARY METABOLISM BY
PHOSPHORYLATION, ON BOTH SER AND THR, THE AFSK GLOBAL REGULATORY
PROTEIN INVOLVED IN THE CONTROL OF SECONDARY METABOLISM.
CC -1- CATALYTIC ACTIVITY: ATP + a protein = ADP + a phosphoprotein.
CC -1- PFM: AUTOPHOSPHORYLATED AT THR AND SER RESIDUES.
CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D45382; BAA08229.2; -.
DR EMBL; AL939120; CAD55483.1; -.
DR PhosSite; P54741; -.
DR InterPro; IPR000719; Prot_kinase.
DR InterPro; IPR002290; Ser_thr_kinase.
DR Pfam; PF00069; Kinase; 1.
DR ProDom; PD000001; Prot_kinase; 1.
DR SMART; SM00564; POK; 9.
DR PROSITE; PS00107; PROTEIN KINASE ATP; 1.
DR PROSITE; PS00108; PROTEIN KINASE ST; 1.
DR PROSITE; PS00111; PROTEIN KINASE DOM; 1.
KW Transferase; Serine/threonine-protein kinase; ATP-binding;
KW phosphorylation; Complete proteome.
FT DOMAIN 16 271
FT NP_BIND 22 30
FT BINDING 44 44
FT ACT_SITE 138 138
FT ACT_SITE 138 138
FT ACT_SITE 138 138
SQ SEQUENCE 799 AA; 83787 MW; 4BB9BED165F6F5B CRC64;

Query Match 4.6%; Score 92; DB 1; Length 799;
Best Local Similarity 24.0%; Pred. No. 67;
Matches 62; Conservative 21; Mismatches 69; Indels 106; Gaps 15;

QY 118 AGORLHMLTGLSAAM-----AATNRMR-----TERMYGDDY--- 154
DB 556 AGAALHD---GTAVWQDARLRALDARTGDERMSYPIGDAASGGVPRVLTQAPDGVYV 612
QY 155 ---VRAVEITHGKGMHVHVALM-----FSGDVSSENILBSFSDAMPDRWT 198
DB 613 AAGTRVLALFVAGVWHEFADAVFLAPTFVPGPAVVGGV---YLADY----- 660
QY 193 SKVLSLGPAPLRNSGSLDYRKIGER---DQYLAAYLTIKASGVMEVSGDGSGRH 254
DB 661 ---LGTVVALATDRDRMRATEARSTDPVL-----VAAG-HVHVGSGKGL----- 704
QY 255 GNPAPWEIADVADGVDQALFMRFEFGS--MGRRAIAMS----- 294
DB 705 -----YTIDAVYGTGK---NR-PQAGSDIYGAPVAGRRIHFGSSDHLITLKADDG 752
QY 295 GLRPAAGLGAELTDQIV 312
DB 753 RLFWKLATGGEITGSPV 770

RESULT 13
PERX BRARA STANDARD; PRT; 296 AA.
ID PERX BRARA 000434;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Peroxidase P7 (EC 1.11.1.7) (TP7).
OS Brassica rapa (turnip).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

```

OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eucosids II; Brassicales; Brassicaceae; Brassica.
 OX NCBI_TaxID=51350;
 RN [1]
 RP SEQUENCE.
 RC STRAIN=cv. Blanc dur d'hiver;
 RX MEDLINE=61003872; PubMed=7408864;
 RA Mazza G., Melinder K.G.;
 RT "Covalent structure of turnip peroxidase 7. Cyanogen bromide
 fragments, complete structure and comparison to horseradish
 peroxidase C.";
 RL Eur. J. Biochem. 108:481-489(1980).
 RN [2]
 RP SEQUENCE OF 32-65 AND 161-175.
 RX MEDLINE=77138218; PubMed=849740;
 RA Melinder K.G., Mazza G.;
 RT "Amino-acid sequences of heme-linked, histidine-containing peptides
 of five peroxidases from horseradish and turnip.";
 RL Eur. J. Biochem. 73:353-358(1977).
 CC -1- FUNCTION: Removal of H(2)O(2), oxidation of toxic reductants,
 biosynthesis and degradation of lignin, suberization, auxin
 catabolism, response to environmental stresses such as wounding,
 pathogen attack and oxidative stress. These functions might be
 dependent on each isozyme/isotform in each plant tissue.
 CC -1- CATALYTIC ACTIVITY: Donor + H(2)O(2) = oxidized donor + 2 H(2)O.
 CC -1- COFACTOR: Binds 1 protoheme IX, 1 iron(III) ion and 2 calcium
 ions.
 CC -1- MISCELLANEOUS: The protein shown, TP7, is the principal
 isoperoxidase during winter in turnip.
 CC -1- SIMILARITY: BELONGS TO THE PEROXIDASE FAMILY. CLASSICAL PLANT
 CC (CLASS III) PEROXIDASE SUBFAMILY.
 CC PIR; A91094; OBNB7.
 DR HSP; P22195; ISCH.
 DR InterPro; IPR002016; Peroxidase.
 DR Pfam; PF00141; peroxidase; 1.
 DR PRINTS; PR00458; PEROXIDASE.
 DR PROSITE; PS00435; PEROXIDASE_1; 1.
 DR PROSITE; PS00436; PEROXIDASE_2; 1.
 DR PROSITE; PS00873; PEROXIDASE_4; 1.
 KM Oxidoreductase; Glycoprotein; Peroxidase; Iron; Heme;
 KM Multigene family; Calcium; Pyroglutamate carboxylic acid.
 FT METAL 43 43 CALCIUM 1 (BY SIMILARITY).
 FT METAL 46 46 CALCIUM 1 (VIA CARBOXYL OXYGEN)
 FT METAL 48 48 CALCIUM 1 (VIA CARBOXYL OXYGEN)
 FT METAL 48 48 CALCIUM 1 (VIA CARBOXYL OXYGEN)
 FT METAL 50 50 CALCIUM 1 (BY SIMILARITY).
 FT METAL 52 52 CALCIUM 1 (BY SIMILARITY).
 FT METAL 170 170 CALCIUM 2 (BY SIMILARITY).
 FT METAL 216 216 CALCIUM 2 (BY SIMILARITY).
 FT METAL 219 219 CALCIUM 2 (BY SIMILARITY).
 FT METAL 224 224 CALCIUM 2 (BY SIMILARITY).
 FT ACT SITE 38 38 BY SIMILARITY.
 FT ACT SITE 42 42 DISTAL HISTIDINE.
 FT ACT SITE 70 70 HYDROGEN-BOUND (BY SIMILARITY).
 FT ACT SITE 139 139 SUBSTRATE-BINDING (BY SIMILARITY).
 FT METAL 169 169 IRON (HEME AXIAL LIGAND).
 FT ACT SITE 241 241 HYDROGEN-BOUND (BY SIMILARITY).
 FT DISULFID 11 91
 FT DISULFID 44 49
 FT DISULFID 97 292
 FT DISULFID 176 201
 FT MOD RES 1 1 PYRROLIDONE CARBOXYLIC ACID.
 FT CARBOHYD 185 185 N-LINKED (GLCNAC. . .).
 SQ SEQUENCE 296 AA; 31086 MW; 53C9CCE59B2A7937 CRC64;

Query Match 4.5%; Score 91; DB 1; Length 296;
 Best Local Similarity 19.1%; Pred. No. 2.6; Mismatches 79; Indels 112; Gaps 15;
 Matches 58; Conservative 55;

QY 16 PVLVSSDKRGIRHET--RPKL-----QQTTFSTFNACGRPSGVN- 54
 DB 12 PNLSTVSKGVKSAVSSQPRMGASILRLFFHDCFVNGCGSILDDTSTSTGQNAQPNR 71

QY 55 ----GTVIVNGPKGS---GFGGLRSG-----GKMWICPCCKAGYGAHA 91
 DB 72 NSARGFTVINDISAVEKACPGVSCADLIALAARDSTVQLGPPNNV-----KVGRDA 126
 QY 92 DEISQVVAHQ-----LGTSVAVVTVWNTATGQ-----RLHDL 125
 DB 127 KTSQAQANANINIPAPMSLSQILSSFAVGLSTRDVALSGAHTIGQSRGVNFARVYN- 185
 QY 126 WTGLSAAW-----KATNGRRKTRREMYGCGYVRAVEI---TGKNGMHVHTA 173
 DB 186 ETININAFATLRGSCPRAGSG-----DANLAPDINSATSPDMSYFQNLMA 233
 QY 174 -----LMPFGDVSENILSFSD--AMPDRWTSKIVSGFAFLRNSGLDYRKIG 222
 DB 234 QRLHSDQVTLFRNGSDTSIVRGYSNPSFNSPFAAMIKGDISPLTSSG-ELRKVC 292
 QY 223 GQAD 226
 DB 293 GKTN 296

RESULT 14
 RECF_MYCPA STANDARD; PRT; 385 AA.
 AC Q9L7L5;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE DNA replication and repair protein recf.
 GN RECF.
 OS Mycobacterium paratuberculosis.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 OX NCBI_TaxID=1770;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Zhang Q., Kapur V.;
 RT "Genomic organization of the Mycobacterium avium subsp.
 paratuberculosis origin of replication region."
 RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: THE RECF PROTEIN IS INVOLVED IN DNA METABOLISM; IT IS
 REQUIRED FOR DNA REPLICATION AND NORMAL SOS INDUCIBILITY. RECF
 BINDS PREFERENTIALLY TO SINGLE-STRANDED, LINEAR DNA. IT ALSO SEEMS
 TO BIND ATP (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (Potential).
 CC -1- SIMILARITY: BELONGS TO THE RECF FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@sib-sib.ch).
 CC EMBL; AF222789; AAF33693.1; --
 DR HAMAP; MF_00365; -; 1.
 DR InterPro; IPR001238; Recf.
 DR InterPro; IPR003395; SMC_N.
 DR Pfam; PF02463; SMC_N; 1.
 DR TIGRFAMs; TIGR00611; recf; 1.
 DR PROSITE; PS00617; RECF_1; 1.
 DR PROSITE; PS00618; RECF_2; 1.
 KM DNA damage; DNA replication; DNA-binding; SOS response; DNA repair;
 KW ATP-binding.
 FT NP BIND 30 37 ATP (POTENTIAL).
 SQ SEQUENCE 385 AA; 42001 MW; 914CBG201971DFP CRC64;

Query Match 4.5%; Score 90; DB 1; Length 385;
 Best Local Similarity 23.3%; Pred. No. 4.3; Mismatches 120; Indels 102; Gaps 17;
 Matches 78; Conservative 35;

```

QY 55 GTTIVNGPKSGFGGLRSCGKGMICPCACAGVGAHRADEISQVVAHQLTGTSVAVMTMTM 114
DB 24 GRATVFIG--SNGFEKTMLEALMWY-----SSTLSHRYGTDAPILR-----AGADRAVAVSTI 73
QY 115 RHTAGQELH-DLMTGLSAAMKAATNGRRMTEREMTGCQGVAAVEITGKNGMHVHVA 173
DB 74 VVNDGRCAVDLEIAGARKANKRLNRSPVSTRBVL-----GVLRVAVFA-----PRD 121
QY 174 LLMFSGDVSEN-----ILSEFSDA----- 192
DB 122 LAIVRGDPSEERRRITUDLATIRRPALAAVADYDKVLKORTALIKLSGARHRSDRGALD 181
QY 193 MEDWTSKLVSLG--FAA--PLNNGGLDVRKIGBADQVLA-----AYLTXTIASG 239
DB 182 TLVDVDSRLAEVGAQLMAARIDLVNQAPVEYK-----AYQLAPGSRRAASIGYRSSLGAA 237
QY 240 VGMVSGDGGK-----SGRHGR-APWEIADAVAGDGPQALEMRPE-----FSSMR 287
DB 238 ASAEVNAAGDRDYELAALLAGLAARDAELERGMCLVGPFRDLEMLGQVAKGFASHGE 297
QY 288 RAIVMSRGLPARAGIAGAEITDAQIVEOEESAPVAV 322
DB 298 ---SWSLALSLR-----LAAYELRLADESDPVIL 323

```

RESULT 15

```

FLB1 CAMJE STANDARD; PRT; 571 AA.
ID FLB1 CAMJE
AC P56964; O9PMW1;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Flagellin B.
DS FLAB OR CU1338C.
OS Campylobacter jejuni.
OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacterales;
OC Campylobacteraceae; Campylobacter.
CX NCBI_TaxID=197;

```

SEQUENCE FROM N.A.

```

RP STRAIN=NCYC 11168;
RX MEDLINE=20150912; PubMed=10688204;
RA Parkhill J., Wren B.W., Mungall K., Kettle J.M., Churcher C.,
RA Basham D., Chillingworth T., Davies R.M., Feltham T., Holtroyd S.,
RA Vagstad K., Karlyshev A.V., Moule S., Pallen M.J., Penn C.W.,
RA Quail M.A., Rajandream M.A., Rutherford K.M., van Vliet A.H.M.,
RA Whitehead S., Barrall B.G.;
RT "The genome sequence of the food-borne pathogen Campylobacter jejuni
RT reveals hypervariable sequences.";
RL Nature 403:665-668(2000)
CC -!- FUNCTION: FLAGELLIN IS THE SUBUNIT PROTEIN WHICH POLYMERIZES TO
CC -!- FORM THE FILAMENTS OF BACTERIAL FLAGELLA.
CC -!- SUBUNIT: HETEROPOLYMER OF FLAA AND FLAB.
CC -!- SIMILARITY: BELONGS TO THE BACTERIAL FLAGELLIN FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC

```

```

DR EMBL; AL139078; CAB73765.1; -
DR PIR; G81277; G81277.
DR InterPro; IPR001029; Flagellin_C.
DR InterPro; IPR001492; Flagellin_N.
DR Pfam; PF00700; Flagellin_C_1.
DR Pfam; PF00669; Flagellin_N_1.
DR PRINTS; PR00207; FLAGELLIN.
DR Flagella; Complete proteome.
KM INIT_MBT 0
SQ SEQUENCE 571 AA; 59054 MW; ACE2D9B300F05053 CRC64;

```

```

Query Match 4.5%; Score 90; DB 1; Length 571;
Best Local Similarity 23.5%; Pred. No. 6.7;
Matches 67; Conservative 28; Mismatches 78; Indels 112; Gaps 16;

QY 20 SDRGIRHLEIRPLQOI-----TSETFNACGRPISGVNGVTVNGPKSGFGGLRSC 73
DB 223 SADKTGVAVATFTVTRGIAAVRAGTTSDTF-----AINGVTI----- 259
QY 74 GKGMICPCACAGVGAHRADEISQVVAHQLTGTSVAVMTMTMTAGQRLHDLMTGLSAAM 133
DB 260 -----GVAVYEDGD-----GNGLVVAALNSVKDT-----TGVEAST 290
QY 134 KAATNGRRMTEREMTGCQGVAAVEITGKNGMHVHVALLMFSGDVSEN-----ILE 187
DB 291 DA--NGQLLLTSRE-----GRGIKI--DNIGGGAIRINA-----DMKENYGRSLVKN 334
QY 188 SPSDAMPDRWTSKLVSLGFAA-----PLNNGG-----LDVRKIGBADQV----- 228
DB 335 DKKDILIS--GSNLSAGFGATQFISQASVLSRESKGRFPANTADAMGFGSANKGVVLGG 392
QY 229 ---LAAYLTK-----IASGVGMEVSGDCKSGRHGRAPWEIAYDA 266
DB 393 YGSVAIYMSAGSGFSSGSGYSVSGKNYGTGPAN---ALATSA 433

```

Search completed: November 8, 2003, 23:20:08
Job time : 26 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 8, 2003, 23:09:28 ; Search time 66 Seconds
(without alignments)
1481.848 Million cell updates/sec

Title: US-10-007-527a-2
Perfect score: 2005
Sequence: 1 MTSVAEHLGKDRPPVLS.....HLHYRLPADYRPPISVRK 379

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

1: SP_ARCHAEA:*
2: SP_BACTERIA:*
3: SP_FUNGI:*
4: SP_HUMAN:*
5: SP_INVERTEBRATE:*
6: SP_MAMMAL:*
7: SP_MHC:*
8: SP_ORGANELLE:*
9: SP_PHAGE:*
10: SP_PLANT:*
11: SP_RODENT:*
12: SP_VIRUS:*
13: SP_VERTEBRATE:*
14: SP_UNCLASSIFIED:*
15: SP_VIRUS:*
16: SP_BACTERIAP:*
17: SP_ARCHAEA:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	609.5	30.4	459	2 P94171	P94171 arcobacte
2	569	28.4	433	2 Q8GDF1	Q8GDF1 propionibac
3	497	24.8	403	2 Q69764	Q69764 propionibac
4	424.5	21.2	476	2 Q9RT00	Q9RT00 streptomyce
5	374.5	18.7	405	9 Q8SCH7	Q8SCH7 propionibac
6	373	18.6	380	2 Q9X3U7	Q9X3U7 bifidobacte
7	360.5	18.0	622	2 Q8GSD7	Q8GSD7 streptomyce
8	345	17.2	471	2 Q8RIU6	Q8RIU6 pseudomonas
9	342.5	17.1	461	2 Q9J141	Q9J141 corynebacte
10	326	16.3	401	2 Q9R881	Q9R881 corynebacte
11	315.5	15.7	312	2 Q8GHI1	Q8GHI1 pseudomonas
12	293.5	14.6	386	2 Q9ACT3	Q9ACT3 rhizobium m
13	277.5	13.8	356	2 Q52220	Q52220 acinetobact
14	266.5	13.3	528	2 Q57477	Q57477 streptomyce
15	265	13.2	478	2 Q8RNS3	Q8RNS3 corynebacte
16	256.5	12.8	505	2 Q83016	Q83016 streptomyce

17	235.5	11.7	481	2 Q54245	Q54245 streptomyce
18	230.5	11.5	437	2 Q52205	Q52205 streptomyce
19	194	9.7	396	2 Q69069	Q69069 streptomyce
20	169	8.4	240	2 Q45282	Q45282 corynebacte
21	161.5	8.1	261	2 Q8G156	Q8G156 corynebacte
22	123.5	6.2	315	2 Q56129	Q56129 streptococc
23	122.5	6.1	315	2 Q9RNV8	Q9RNV8 streptococc
24	122.5	6.1	315	2 Q9RNM4	Q9RNM4 streptococc
25	121.5	6.1	315	2 Q8RNM5	Q8RNM5 streptococc
26	119	5.9	319	2 Q9ZNC0	Q9ZNC0 nitrosomona
27	115	5.7	336	2 Q06460	Q06460 synchocyst
28	114.5	5.7	315	2 Q9X3N4	Q9X3N4 streptococc
29	114	5.7	314	2 Q30850	Q30850 streptococc
30	111.5	5.6	276	2 Q9K4S9	Q9K4S9 pseudomonas
31	111.5	5.6	320	2 Q52769	Q52769 streptococc
32	111.5	5.6	320	2 Q54265	Q54265 streptococc
33	111.5	5.6	320	2 Q52776	Q52776 streptococc
34	111.5	5.6	333	2 Q00452	Q00452 bacillus su
35	111	5.5	329	2 Q9F7Y4	Q9F7Y4 salmonella
36	110.5	5.5	339	2 Q45447	Q45447 bacillus su
37	109.5	5.5	315	2 Q9X9N2	Q9X9N2 streptococc
38	108.5	5.4	314	2 Q9X9C2	Q9X9C2 streptococc
39	107.5	5.4	321	2 Q99135	Q99135 shigella so
40	105.5	5.3	743	13 Q9PUT1	Q9PUT1 gallus gall
41	105	5.2	381	16 Q915P2	Q915P2 pseudomonas
42	104.5	5.2	340	2 Q45450	Q45450 bacillus su
43	104.5	5.2	340	2 Q9X3Y9	Q9X3Y9 bacillus su
44	104.5	5.2	988	5 Q17434	Q17434 nephila cla
45	104	5.2	313	2 Q9X9P2	Q9X9P2 streptococc

ALIGNMENTS

RESULT 1
ID P94171 PRELIMINARY; PRT; 459 AA.
AC P94171;
DT 01-MAY-1997 (TREMBLrel. 03, Created)
DR 01-MAY-1997 (TREMBLrel. 03, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE Rep.
GN Rep.
KM Rep.
OS Arcanobacterium pyogenes.
OG Plasmid PAP1.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Actinomycetales; Actinomycetaceae; Arcanobacterium.
OX NCBI_TaxID=1661;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BBR1;
RX MEDLINE=98292760; PubMed=9620977;
RA Billington S.J., Jost B.H., Songer J.G.;
RT "The Arcanobacterium (Actinomycetes) pyogenes plasmid PAP1 is a member
of the pUT101/pUT1 family of rolling circle replication plasmids.";
RL J. Bacteriol. 180:3233-3236 (1998).
DR EMBL; U83788; AAC46399.1; -
DR InterPro; IPR000989; Rep.
DR Pfam; PF01446; Rep; 1.
KW Plasmid.
SQ SEQUENCE 459 AA; 51264 MW; 2A0A4B0A43BF65 CRC64;

Query Match 30.4%; Score 609.5; DB 2; Length 459;
Best Local Similarity 36.3%; Pred. No. 1.2e-38;
Matches 148; Conservative 62; Mismatches 133; Indels 65; Gaps 15;
QY 5 SAEHLGKDRPPVLS-----KRGIRHELRPLQQTTSSETFNACGR-PIGNGV 56
DB SATHPLENTVLTTPVSVESKRTAKRSERYELRDGLAISTESVRKCGRVAPVLSL 126
QY 57 TTVNGPFGSGFGGIRSCGKGMICPCGAGVGAHRADEISQVVAHQGTG-SVAVWTMR 115
DB 127 FAKSDGKAGAGYGGIHTCGSVWACPVCSAKTAAARKTDLQGVNDAVHAGWTVMULTTOR 186

QY 116 HTAGORLHDLMTGLSAAMKATNGRRTREMYGCDYVRAVELTHGKNGHVVHALL 175
 DB 187 HHGQGLKHLMDLSTANMRVTSGRNIEKQFGLVGVRAVELTHGKNGHVVHALL 246
 QY 176 MFGDVSENILSEF-----SDAMPDRTSKLVSGFAAPLRNSG 215
 DB 247 ISEKD---PLTSTFVYQKQGRRLPYPEIYWSDFIAERWAGLAKHGVDF-LRDSG 302
 QY 216 L-----DVRKIGSEADQVLAALTKL---AGVGMEVSSGSGKSRHNRAPRIADA 266
 DB 303 LDWTVAKDADAIN-----YVSKQGTSTDAISEVTLGGFKAKNGNRTPQIAD 354
 QY 267 VG-GDPOALELMEFEFGSKGRRAIAMSRLAPRAGIETDAQIVEQESGAPVMAII 325
 DB 355 LSLGDDVDDKLMEYERKASFGRRALTMKSLRPMALGVQSGDELA-SEEGDADALF 413
 QY 326 PANSMMIRTCADYVFG--EILGLVEAGA-----TWENIRDLHYRLP 366
 DB 414 THDAMQVRR-----FGAAELLDVTESGRAAAYFWLDERR-IDWSLP 455

RESULT 2

Q8GDF1 PRELIMINARY; PRT; 433 AA.
 AC Q8GDF1;
 DT 01-MAR-2003 (TEMBLrel. 23, Created)
 DT 01-MAR-2003 (TEMBLrel. 23, Last sequence update)
 DT 01-MAR-2003 (TEMBLrel. 23, Last annotation update)
 DE REP.
 GN REP.
 OS Propionibacterium granulosum.
 OC Plasmid cryptic plasmid pPG01.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Propionibacteriaceae; Propionibacteriaceae; Propionibacterium.
 RX NCBI_TaxID=33011;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=PF283;
 RA Farrar M.D., Holland K.T.;
 RT "Isolation and characterisation of a cryptic plasmid from the human
 RT skin commensal Propionibacterium granulosum";
 RL Submitted (SERP-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AY150274; AN78123.1; -.
 KW Plasmid.
 SQ SEQUENCE 433 AA; 48340 MW; 1ABG34COFBF8678F CRC64;

Query Match 28.4%; Score 569; DB 2; Length 433;
 Best Local Similarity 35.6%; Pred. No. 1.4e-35;
 Matches 136; Conservative 63; Mismatches 145; Indels 38; Gaps 14;

QY 16 PVLVSDKRGIRHELRPKQ-QITSETFNACGR-PISGVNGVTIVNGPKSGFGGLRS 72
 DB 51 PCRKPNEGRRHREYEMRGLLENPQVMEPLERYKCGAVPV--GRILAMGHHGGYAGLAT 108
 QY 73 CGKGMVCCGAGVGAHRADEISOVAHOLGCG-STAVMTMTMRHAGORLHDLMTGLSA 131
 DB 109 CGSWACPVCAKISARREDELARVQVAVGLGPKVSMILTITRHHAGDLELWASLOS 168
 QY 132 AMXAATNGRRWTEREMYGCDYVRAVELTHGKNGHVVHALLMFGSD---VSENI--- 185
 DB 169 GMAVVTISGRNIEKQFGLVGVRAVELTHGKNGHVVHALLVISKDPTSDTIRRR 228
 QY 186 -----LESFSDAMPDRTSKLVSG--FAAPLRNSGGLDVRKIGSEADQVLA 230
 DB 229 RKQGRRTPYPEEYQRPEDFIARWSRGLRKGVDFA--GGGGLDQWQTAOSGDEALG 285
 QY 231 AYLUKIAS--GVGMEVSSGSGKSRHNRAPRIADAVG--GDPOALELMEFEFGSG 286
 DB 286 RYAKAMSSVDGLANETLGGFKKARGNTFPQILEDPLDSSSTDLIMRYVYASNG 345
 QY 287 RRAIWSRGLPABAGIETDAQIVEQESGAPVMAIIPANSMMIRTC-ADYVFGEL- 344

DB 346 RKALTWSSKGLRDVAGNESEMSDQVAAOQCGEA--VALFDHDMRQIRTAGAFLIDLE 404
 QY 345 LGLVEAGATWENIRDLHYRLP 366
 DB 405 LHSGEGVYAMLKRR-RHYELP 425

RESULT 3

Q69764 PRELIMINARY; PRT; 403 AA.
 AC Q69764;
 DT 01-AUG-1998 (TEMBLrel. 07, Created)
 DT 01-AUG-1998 (TEMBLrel. 07, Last sequence update)
 DT 01-DEC-2001 (TEMBLrel. 19, Last annotation update)
 DE Putative rep protein.
 GN REP.
 OS Propionibacterium freudenreichii.
 OC Plasmid pUMET108.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Propionibacteriaceae; Propionibacteriaceae; Propionibacterium.
 RX NCBI_TaxID=1744;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Dasein G.H., Miescher S., Teuber M., Meile L.;
 RT "Molecular analysis of plasmid pUMET108, a plasmid isolated from
 RT Propionibacterium freudenreichii";
 RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AJ006662; CAA07175.1; -.
 DR InterPro; IPR000989; Rep.
 DR Pfam; PF01446; Rep. 1.
 KW Plasmid.
 SQ SEQUENCE 403 AA; 43502 MW; 340F40AA697E880C CRC64;

Query Match 24.8%; Score 497; DB 2; Length 403;
 Best Local Similarity 34.0%; Pred. No. 4.5e-30;
 Matches 127; Conservative 63; Mismatches 144; Indels 40; Gaps 15;

QY 19 VSSDKRGIRHELRPKQ-QITSETFNACGR-PISGVNGVTIVNGPKSGFGGLRS 77
 DB 47 VAADKRGHRSRYVWLRHTSLKRVACGVAAASVAVGRCSDRAGPAGIQSGSVW 106
 QY 78 ICPCCAGVGAHRADEISQV---AHQLGTSVAVMTMTMRHAGORLHDLMTGLSAW 133
 DB 107 ACPVCNAKIRARGLIEGAAYETWTK--GKVAFTFTVRSRSDSLTAVVDVYASGW 163
 QY 134 KAATNGRRWTEREMYGCDYVRAVELTHGKNGHVVHALLMFGSDVSENILSEPSDAW 193
 DB 164 RRVTSCKGWTSDQLRHGVGVFVRVVEYTHGRNGHVVHALLVPLVGDGPDAL--ALHRM 221
 QY 194 FDRWTSVLVSGFAAPLRNSGGLDVRKIGSEAD-QVLAALTK--IASVGMEVSSGDDK 250
 DB 222 FGRWREGLVLAGLGTPIARA--QDVQMSAATGLDILARTLSAOFQKIGHLTVSQSK 279
 QY 251 SGRH--GNRAPEIADAVAGDPOALELMEFEFGSKGRRAIAMSRLAPRAGIETDA 308
 DB 280 TARSALSTRGWELTGDAAAGLAKKEVGLMEWEKSGRGIQMSAGLRRLGLMVEBD 339
 QY 309 AQIYEQE--ESAPVVAIIPARSM-----MRTCAPYVFGELGVENG-ATWEN 356
 DB 340 DIVIAEEVGSVADVALITDGDWRRLVGOQKLYELRSC-----ELGG--QAGLARM-- 389
 QY 357 LRDH-LHYRLPAD 369
 DB 390 LQHHGIRHELVEVD 403

RESULT 4

Q9RI00 PRELIMINARY; PRT; 476 AA.
 AC Q9RI00;
 DT 01-MAY-2000 (TEMBLrel. 13, Created)
 DT 01-MAY-2000 (TEMBLrel. 13, Last sequence update)
 DT 01-JUN-2001 (TEMBLrel. 17, Last annotation update)

DE Replication protein.
 GN REP.
 OS Streptomyces natalensis.
 OG Plasmid pSN1.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Streptomyces; Streptomycetaceae; Streptomyces.
 OX NCBI_TaxID=68242;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Mendes M.V., Aparicio J.F., Martin J.F.;
 RT "Complete nucleotide sequence and characterization of pSN1 from the
 RT plasmid-producing Streptomyces natalensis that replicates by a
 RT rolling circle mechanism."
 RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AJ243257; CAB62261.1; -.
 DR InterPro; IPR000989; Rep.
 DR Pfam; PF01446; Rep; 1.
 KW Plasmid.
 SQ SEQUENCE 476 AA; 51890 MW; DEB4CB3CC2B5B95 CRC64;
 Query Match 21.2%; Score 424.5; DB 2; Length 476;
 Best Local Similarity 29.7%; Pred. No. 2.1e-24;
 Matches 121; Conservative 53; Mismatches 148; Indels 85; Gaps 14;
 QY 37 ITTSETFNACGREPISGVNGVTTVNGFKSGF--GGLRSCGKMWICPCGAKYGAHRADEIS 95
 Db 1 MTTDKALSRGRALG--GGVTPKVTGPGTAYLAGLATCGKVAHCPCGAKIRSATVELQ 59
 QY 96 QV-----VAHQGTGSVAMVTMTMERT 117
 Db 60 AAGTAMEDIGGNGPKIRPAAAKKDINDRTAEADMALEEDAGLQAGGLGKLTITMHHY 119
 QY 118 AGQRLDLMTGLSAAMKAATN---GRMTEREMVCGDGYVAIVEITGK--NGMHVHA 173
 Db 120 SRHTLABIVTQGDANKKALGQAGDWRKAKKDGVGVFVAMVETVGEANQMPHHV 179
 QY 174 LMFSGDVSENIIESFSDAMFDRWTSKLVSLGPAAPLRNSG--GLDVKRIQGEADQVLAAY 232
 Db 180 LVFEDPKLPEQGDALAEVLYEAMSTALQDVAGYVDFREHGVRLDLSGHEGSGPR--ARY 237
 QY 233 LVTIASG-----VGMVEVGSDGSGRGNRAPEWIA-----VDAVGAD--PQALELREFE 281
 Db 238 LMYDQDGKAAMTAAETKRTDTKAGDGRTPFELIRVLTEDDADDBRAQVYRMQYE 297
 QY 282 FSGMGRRAIAMSRLRPAAGLGAEL--TDAQIVQGESAPVMAI1-----PA 327
 Db 298 TAARGKRALYWSNGRLAALVELDTRDGEIAAEPRQGEALAVIADPWHQHIAARRKG 357
 QY 328 RSMNMTRTCAFYVFGELIGLVEA--GATWENLDHLYRPAADVPP 373
 Db 358 RSLQLKAAEKGGQDQKRALVBSWGLV-----GRDVLPP 392
 RESULT 5
 08SCH7 PRELIMINARY; PRT; 405 AA.
 ID 08SCH7; PRT; 405 AA.
 AC 08SCH7; PRT; 405 AA.
 DT 01-JUN-2002 (TReMBLrel. 21, Created)
 DT 01-JUN-2002 (TReMBLrel. 21, Last sequence update)
 DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
 DE 08SCH7.
 OS Propionibacterium phage phib5.
 OC Viruses; ssDNA viruses; Inoviridae; Inovirus.
 OX NCBI_TaxID=189836;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA MEDLINE=21886396; PubMed=11889111;
 RA Chopin M.C., Rouault A., Ehrlich S.D., Gautier M.;
 RT "Filamentous Phage Active on the Gram-Positive Bacterium
 RT Propionibacterium freudenreichii."
 RL J. Bacteriol. 184:2030-2033(2002).
 RN [2]
 RP SEQUENCE FROM N.A.

RA Chopin M.C., Rouault A., Gautier M.;
 RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF428260; AAL91702.1; -.
 SQ SEQUENCE 405 AA; 44762 MW; DE4B5BC4819D4F90 CRC64;
 Query Match 18.7%; Score 374.5; DB 9; Length 405;
 Best Local Similarity 29.4%; Pred. No. 1.2e-20;
 Matches 111; Conservative 62; Mismatches 154; Indels 51; Gaps 14;
 QY 8 HUSGDRPEVLVSSDKGRHRLPKLQITSETFNACGR--PISGVNGVTTV--NGPK 63
 Db 14 HADGSNR-----RRRVMSARRTLWTGSLQGRVRCGHPSPSLSSGVAIMVRDOST 65
 QY 64 G-----SGFGLSGCKGKMWICPCGAKYGAHRADEISQVVA--HQGTGSVAMVTMTMERT 117
 Db 66 GNVTDAGISGLQHCNSWACPVCAQIISQRMPLSEVLSRHHARG--GRAMTITLVHHD 124
 QY 118 AGQRLDLMTGLSAAMKAATNGRRMTEREMVCGD---VVAIVEITGKNQMHVHAAL 174
 Db 125 RTQSLKTVMDAVAKGMSKATNGRSWDVLSLFGVDGRLPMLRFEVTHGSGMHVHAAL 184
 QY 175 LMFSGDV---SENIIESFSDAMFDRWTSKLVSLGPAAPLRN--SGGLDVKRIQGEADQVLA 230
 Db 185 VLLGEGARREADAVEQIPARWWSMRAVVRQGLKSLSEAHILTEPDAGDGLGILA 244
 QY 231 AYLTKIASGVGMVEVGSDGSGRGNRAPEWIAVDAVGAD-----PQALEL 276
 Db 245 AYFSKAAFEILT---GATKAAQGDNCITPFI--LTLGLDNGEVLRLSGKERTRLMV 300
 QY 277 WREFSGMGRRAIAMSRLRPAAGLGAELT--DQIVQGESAPVMAI1PARWMMIR 334
 Db 301 WHEWEEFSSGRRMGYSKGLFDMILGLDSWTPSDDEILBDNDLDELFTVGHHEWTALR 360
 QY 335 TCAPYVFGELIGLVEAGA 352
 Db 361 D-----RGLADLVLEAVA 373
 RESULT 6
 09X3U7 PRELIMINARY; PRT; 380 AA.
 ID 09X3U7; PRT; 380 AA.
 AC 09X3U7; PRT; 380 AA.
 DT 01-NOV-1999 (TReMBLrel. 12, Created)
 DT 01-NOV-1999 (TReMBLrel. 12, Last sequence update)
 DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)
 DE Replication protein.
 GN REP.
 OS Bifidobacterium breve.
 OC Bacteria; Actinobacteria; Actinobacteridae; Bifidobacteriales;
 OC Bifidobacteriaceae; Bifidobacterium.
 OX NCBI_TaxID=1685;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA STRAIN=NCFB 2258;
 RA MEDLINE=99271179; PubMed=10339821;
 RA O'Riordan K., Fitzgerald G.F.;
 RT "Molecular characterization of a 5.75-kb cryptic plasmid from
 RT Bifidobacterium breve NCFB 2258 and determination of mode of
 RT replication."
 RL FEMS Microbiol. Lett. 174:285-294(1999).
 DR EMBL; AF085719; AAD34709.1; -.
 DR InterPro; IPR000989; Rep.
 DR Pfam; PF01446; Rep; 1.
 KW Plasmid.
 SQ SEQUENCE 380 AA; 42758 MW; D26F59753B604FA CRC64;
 Query Match 18.6%; Score 373; DB 2; Length 380;
 Best Local Similarity 27.9%; Pred. No. 1.4e-20;
 Matches 97; Conservative 56; Mismatches 143; Indels 52; Gaps 11;
 QY 58 IVNGPKSGRGGRGRLSGKGMWICPCGAKYGAHRADEISQVVAHQ----- 102

Db 1 MTNGEK-ARFTGMLGSIWACPTCSAIIIRHRAHVALAIGNHAEKLRKAADQWQH 59

Qy 103 -----GTGSAVAVTMTMHTAGQRLHDIWTGLSAAMKATNGRRRTREMY 149

Db 60 EGQRLPELWVSDSFQYIFGTLTHHDTMPLAMTLDAILKQMTMINGSFWQASBRW 119

Qy 150 GCGYVRAVETITGKGMVHVHALLMFGDVSENLEFSFSDAMDRWTS--KYSLSGF- 206

Db 120 KIRGFVRALEITYGVNGMHPHIFVFLDGDGQREMGQMLDRMTKWKRVAKAYK 179

Qy 207 -----AAPLNSGGLDVR-KIGGADQVLAAYLTIKIASVG-----MEVSGDGKSG 252

Db 180 KQGNPNVAP-NDEHIGIDLQFSGKDAAGTAAAEYITTKIQDGKGVTLTAQETARDING 238

Qy 253 KHGNRAWEIADVAVG--DPQALMREPEFSGMRRALIAMSGLRARAGIAGLTDAQ 310

Db 239 RMSGVNPFQILDGCGGLSDPREDIMLEWYQALRRRCITSRGKEDMEV-ESLEDEE 297

Qy 311 IVEQESAPVMA-IIPARSMWIRTCAPVFEIGLVEAGATWNL 357

Db 298 LAEKDELPGLVGVVFNRYKDIRKABETLADLDAER-EDMQEV 344

RESULT 7

Q8GGD7 PRELIMINARY; PRT; 622 AA.

AC Q8GGD7

DT 01-MAR-2003 (TREMBlrel. 23, Created)

DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)

DE Putative replicase.

OS Streptomyces sp. EN27.

OC Plasmid pEN2701.

OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;

OC Streptomycetaceae; Streptomycetaceae; Streptomycetes.

OX NCBI_TaxID=211464;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=EN27;

RA Coombs J.T., Franco C.M.M., Loria R.;

RT "Complete sequencing and analysis of pEN2701, a novel 13kb plasmid

RT from an endophytic Streptomyces sp.";

RL Plasmid 0:0-0(2002).

DR EMBL; AF533985; AAN76289.1; -.

KW Plasmid.

SQ SEQUENCE 622 AA; 68210 MW; BD467DCEFA209EC95 CRC64;

Query Match

Best Local Similarity 18.0%; Score 360.5; DB 2; Length 622;

Matches 113; Conservative 56; Mismatches 155; Indels 137; Gaps 15;

Qy 24 RGRHRLRPLOQITTEFNACRPISG-VNGVTIVNGPKSGFGGLRSCGKMICPC 82

Db 7 RDRRRRLREAIQRLTSLGIRGCGTGTGELISVH--GTAVRPAGLQSGHSAACPTC 64

Qy 83 AGVGAHRADEISQVVAHQLGTS-SVAMVTMTKH-----TGQ----- 120

Db 65 QARIRARAVELERRALAWLKAGHGYMATLTPHEHVRLASRVGDHKAGQSSCTCP 124

Qy 121 -----RLHDLMTGLS-----AAMKA 135

Db 125 CKCPPLAPGQKRPCTCACPVDPQOLARVVAAMGVGGTMMWGRDVIQDGAAPAPFA 184

Qy 136 A-----TNGRRWTE-----REMYGCGGVRAVEITGKNGM 168

Db 185 TAVDRLLALETGKRYVRRGEPDRSPGPAVWQEGFRDWRGIVGTRTITITWANGM 244

Qy 169 VHVHALLMFGDVSENLEFSFSDAMDRWTSKIVSGFAAPLN-----SGGLDV 218

Db 245 SHAHVLWTEDEADLRAEKEIEBELTNRMAKRCAGVGLTPPAGELRPDGRIGKGVAV 304

Qy 219 RKIGGADQVLAAYLTIKIASG--VGMVEVSGDK--SGHGNRAPEIADVAVGDPQAL 274

Db 305 TAATREKAGVGVYVTKIQEGGNLAMEMTRADLTKARQKXKA-LELAAVAASGNEAL 363

Qy 275 ELWREPEFSGMRRALIAMSGLRARAGIAGLTDAQIYQESAPVMAIIPARSMWIR 334

Db 364 DLWHEPEFATRGQCLTWSQGLRERL--AEIVELDNREDRIPEEFADTEEPSL--- 417

Qy 335 TCAPYVFEIGLVEAGATWNLRLHRLRAADVPRPII 375

Db 418 -----LITRAAMKSI-----VPYGARPDIL 439

RESULT 8

Q8RTU6 PRELIMINARY; PRT; 471 AA.

AC Q8RTU6

DT 01-JUN-2002 (TREMBlrel. 21, Created)

DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)

DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)

DE Hypothetical Rep protein.

GN REP.

OS Pseudomonas fulva.

OC Plasmid pN110.

OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;

OC Pseudomonadaceae; Pseudomonas.

OX NCBI_TaxID=47880;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=IF-4.

RA Itoh N., Kawanami T., Nitta C., Iwata N., Usami S., Abe Y., Koide Y.;

RT "Complete nucleotide sequence of pN110 plasmid in Pseudomonas, the

RT region required for its replication, and the construction of improved

RT Escherichia and Pseudomonas shuttle vector pUK73.";

RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.

DR EMBL; AB084167; BAB90984.1; -.

DR InterPro; IPR000989; Rep.

DR Pfam; PF01446; Rep.1.

SW Plasmid; Hypothetical protein.

SQ SEQUENCE 471 AA; 51989 MW; C44438B8BFEFE1941 CRC64;

Query Match

Best Local Similarity 17.2%; Score 345; DB 2; Length 471;

Matches 89; Conservative 57; Mismatches 147; Indels 40; Gaps 9;

Qy 49 PISGVNGVTIVNGPKSGFGGLRSCGKMICPCAGKVAHRADEISQVVAHQLGTS-SV 107

Db 76 PVAGMNVHVASRHSFAFYSGVMNCGSVNACPVCAKVOBRREVEYAKAVTWADAQGLQ 135

Qy 108 AMTMTMRTHTAGQRLHDIWTGLSAAMKATNGRRRTREMYGCGGVRAVEITGKNGM 167

Db 136 AMVTLPFPHAYMOQLVLEQKTKALKYLRSGSPWTRFKATGYQGLIRSLTTHQNGM 195

Qy 168 VHVHALLMFGDVSENLEFSFSDAMDRWTSKIVSGFAAPLN-----TGQ----- 227

Db 196 HPHTHLMFADAGDADDTMK--TVLERNKISCARAGL-----IDL-----DNAQ 239

Qy 228 VLA-----AVLTIKIAS-----GVMEVSGDGKSGRHNRAPEIADVAVG 268

Db 240 VAAFEANAIDVKGCTASDYLAKODSRHNGVDAELAKASTKAGRAKGFALTLTLPD 299

Qy 269 GDPQALEMREPEFSGMRRALIAMSGLRARAGIAGLTDAQIYQESAPVMAIIPAR 328

Db 300 GDRAGHRLFDYAAAMKGRQLSAGLAKAGV-IGQSDVEVAEQRDDADLGHILE 358

Qy 329 SMNMIRTCAPVFEIGLVEAGATWNLRLH 361

Db 359 QMKLIRQAG--LRAKVIDLAESNGVVAIQ-HL 388

RESULT 9

Q93L41 PRELIMINARY; PRT; 461 AA.

AC Q93L41

DT 01-DEC-2001 (TREMBlrel. 19, Created)

DT 01-DEC-2001 (Tremblrel. 19, last sequence update)
 DT 01-JUN-2002 (Tremblrel. 21, last annotation update)
 DE Putative replication initiation protein.
 GN REP.
 OS Corynebacterium callunae.
 OC Plasmid pCC1.
 CC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 CC Corynebacterineae; Corynebacteriaceae; Corynebacterium.
 NCBI_TaxID=1721;
 RN [1]
 RN SEQUENCE FROM N.A.
 RC STRAIN=ATCC 15911;
 RA Venkova T., Patek M., Neaveva J.;
 RT "Plasmid pCC1 from Corynebacterium callunae";
 RI Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AJ308231; CAC59945.1; -;
 DR InterPro; IPR000989; Rep.
 DR Pfam; PF01446; Rep; 1.
 KW Plasmid.
 SQ SEQUENCE 461 AA; 50964 MW; 8747737EC2CB0701 CRC64;
 Query Match 17.1%; Score 342.5; DB 2; Length 461;
 Best Local Similarity 28.2%; Pred. No. 4.1e-18;
 Matches 102; Conservative 50; Mismatches 163; Indels 47; Gaps 12;
 QY 29 ELRPKQLQITTFEPNACGRPISGVN-GVTIVNGPKSG-FQGLRSCGKMWICPCAGKV 86
 DB ERQKAMWITNIALGCHRMRSKSGSLKANNNGSRFAGLQNSNVWASPIPAVAI 125
 QY 87 GARRADEISQVVA-HQL-GTGSVAVMTMTMTHTAGORLHDMTGLSAWKAATNGRWR 143
 DB 126 GKRVRVQATTAATVKNMMDHNASVLLTLTLPLHGTALAKDLALKAGWAGITTSWK 185
 QY 144 TEREMGCDYVRAVEITGKNGHVHVALMFSGDVSENIIESFSDAMFDRWTKIVS 203
 DB 186 KDRVAYRLPMWMAKALITTHQNSFHFVLFCEKRTSTTEVVALKALFPRYAKLEK 245
 QY 204 LGRAPLRNSGGIDVVKIGLEADQV-LAAYLTK-1AS-GVGMVSGDGSGRGHNAAP 259
 DB 246 HGQRPSWEH-GIDLVQSTGRDAIMWGAITAGLAESWNAASEVAGQAFKAKGTNTP 304
 QY 260 METAVDAVGDDPOLA-----LMREFEFGSMGRRAIAMSRLARAGIAGELTDAQIV- 312
 DB 305 WQILDIDIAAGAPGPRRVAALMREYATRTGKTSSTGAKALSTNV-LKDEYAT 363
 QY 313 -----ESESAPVNVAIIPARSWMITRTCAPYVEGELGLVAGATWENLRDHLHPLPA 367
 DB 364 GEVLGEDEATEDYVADIPAKA-----MAELCDVHKRLDI 399
 QY 368 AD 369
 DB 400 AN 401
 RESULT 10
 ID Q9R881 PRELIMINARY; PRT; 401 AA.
 AC Q9R881;
 DT 01-MAY-2000 (Tremblrel. 13, Created)
 DT 01-MAY-2000 (Tremblrel. 13, last sequence update)
 DT 01-MAY-2000 (Tremblrel. 13, last annotation update)
 DE Replicase.
 GN REPBL1.
 OS Corynebacterium glutamicum (Brevibacterium flavum).
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 CC Corynebacterineae; Corynebacteriaceae; Corynebacterium.
 NCBI_TaxID=1718;
 RN [1]
 RN SEQUENCE FROM N.A.
 RC STRAIN=ATCC 21086;
 RX MEDLINE=96276208; Pubmed=8693028;
 AKRI S., Reyes O., Leblon G.;

RT "Electrotransformation of highly DNA-restrictive corynebacteria with
 RT synthetic DNA"; (1996).
 RU Plasmid 35:62-66 (1996).
 RN [2]
 RN SEQUENCE FROM N.A.
 RC STRAIN=ATCC 21086;
 RA Favay S., Leblon G., Reyes O.;
 RT "Replication of the Corynebacterium glutamicum ATCC 21086 plasmid
 RT pbl1";
 RU Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RN SEQUENCE FROM N.A.
 RC STRAIN=ATCC 21086;
 RA Akri S., Reyes O., Leblon G.;
 RI Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF092037; AAD08690.1; -;
 KW Plasmid.
 SQ SEQUENCE 401 AA; 44481 MW; 5E0590D936132C4D CRC64;
 Query Match 16.3%; Score 326; DB 2; Length 401;
 Best Local Similarity 29.1%; Pred. No. 6.3e-17;
 Matches 98; Conservative 58; Mismatches 131; Indels 50; Gaps 15;
 QY 34 LQGITTFEPNACGR-----PIGVNGVTIVNGPKSGFQGLRSCGKMWICPCAGKV 86
 DB 1 MYKITSKALAGCHRRRDEAVAVSWSS-----NG-ASQFGLQNSHRWGSPLALEV 53
 QY 87 GARRADEISQVVAHQLGTGSVAVMTM-TMHTTAGORLHDMTGLSAWKAATNGRWRTE 145
 DB 54 MGRRELEIATNTGNHLAGALMMPVGYRHRSGSFQAVEGICITATSSWTKSQWKE 113
 QY 146 REMYGCDYVRAVEITGKNGHVHVALMFSGDVSENIIESFSDAMFDRWTKIVS 204
 DB 114 RAYGVGHTYSDYEIVDSWANGWHLRNMILLPLDPLSDDEKAFEDSWFSWAGVKA 173
 QY 205 GFAPLRNSG-GHD-VRTKIGRADQVLAAYLTKASGVGMVSGDDGSGRGHNAAPKEI 262
 DB 174 GMDAPLRNSGKIDQVSTGSDAK-WATYLLK--GMSQELTGSATKTAGSGSTYPRQW 229
 QY 263 AVDAVGDDPOLA-----LMREFEFGSMGRRAIAMSRLARAGIAGELTDAQIV- 311
 DB 230 -LDMADQSDAGDMDAVLVARNREYEVGSKNLRS-SNGRG--AKRALGIDVIDADVARE 285
 QY 312 VEGE-----ESAPVNVAIIPARSWMITRT 335
 DB 286 MEELYKLAGLEAPERVESTVAVALVXDDMKLQIS 322
 RESULT 11
 ID Q8GHP1 PRELIMINARY; PRT; 312 AA.
 AC Q8GHP1;
 DT 01-MAR-2003 (Tremblrel. 23, Created)
 DT 01-MAR-2003 (Tremblrel. 23, last sequence update)
 DT 01-MAR-2003 (Tremblrel. 23, last annotation update)
 DE Hypothetical protein.
 OS Pseudomonas putida.
 CC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
 CC Pseudomonadaceae; Pseudomonas.
 NCBI_TaxID=303;
 RN [1]
 RN SEQUENCE FROM N.A.
 RA Yu Q., Zhang J., Ding Z., Yu J., Huang D.;
 RI Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF273219; AA013751.1; -;
 KW Hypothetical protein; Plasmid.
 SQ SEQUENCE 312 AA; 34777 MW; 56FBAD068984B982 CRC64;
 Query Match 15.7%; Score 315.5; DB 2; Length 312;
 Best Local Similarity 27.8%; Pred. No. 2.9e-16;
 Matches 80; Conservative 45; Mismatches 124; Indels 39; Gaps 8;

QY 72 SCGGKMTCPCCAGKVGANRADEISQVVAHQLTGTG--SVANVTMTMTTACQRLHDLMTGLS 130
 DB 2 NCGSVMACPVCAAAVQRRREVEAKAVAMDAQGLQAAVTLTPHYAQDLAEILLQOQK 61
 QY 131 AAMRAATNGRMRTTEREMYGCDGVYRAVEITTHGKNGHVVHALLMFGSDVSENITLES 190
 DB 62 TALKYLABEGSWTETFKATGYOGLIRSELTHQNGMHPHTEHLMVDACTADIMTK-- 119
 QY 191 DAMFDRMTSKLVISGFAPLRNSGGLDVRRKIGEBADQVLA-----AYLYK 235
 DB 120 -TVLERWKTSCARAGL-----LDLNNV-----DQVAFAFAHADVKGWCTASPYLAK 165
 QY 236 IAS-----GVGMEVSGSGKSGRHGNRAPEIADVAGDPQALMREFEFGSMGRRAIA 291
 DB 166 QDDSRHMGVDAEIKASTKAGRAKGRPALALPDDGDKAGHRLDYAAAMKGRQLP 225
 QY 292 WSRGLRARAAGLGAELTDQIVOEESAPVVAIIPARSMWMTRT--CA 337
 DB 226 WSAGLKAKVGV-IEQSDBEVAEBQDDADLLGHLIEIOMKLIHQAGCA 272

RESULT 12

QY 9AGT3 PRELIMINARY; PRT; 396 AA.
 AC Q9AGT3;
 DT 01-JUN-2001 (Tremblrel. 17, Created)
 DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)
 DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
 DE Rep protein.
 OS Rhizobium meliloti (Sinorhizobium meliloti).
 OG Plasmid pRm1132f.
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
 OC Rhizobiaceae; Sinorhizobium.
 CX NCBI_Taxid=382;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=21172893; PubMed=11274136;
 RA Barran L.R., Ritchot N., Bromfield B.S.P.;
 RT "Sinorhizobium meliloti Plasmid pRm1132f Replicates by a Rolling-
 Circle Mechanism."
 RL J. Bacteriol. 183:2704-2708(2001).
 DR EMBL; AF327371; AAG59887.1; -.
 KW Plasmid.
 SQ SEQUENCE 396 AA; 42948 MW; 89EA71F998579FE7 CRC64;

Query Match 14.6%; Score 293.5; DB 2; Length 396;
 Best Local Similarity 29.4%; Pred. No. 2e-14;
 Matches 77; Conservative 47; Mismatches 117; Indels 21; Gaps 8;

QY 64 GSGFGGLRSCGKMTCPCCAGKVGANRADEISQVVAHQLTGTG--SVANVTMTMTTACQRLHDLMTGLS 121
 DB 9 GARTSGTFPACDSGMLCVCAAPRKQERDRAEVEFHVKAFRDQGMVCTLVVRSRGG 68
 QY 122 LHDIMTGLSAAMKATNGRMRTTEREMYGCDGVYRAVEITTHGK--NMHVVHALLMFGSD 180
 DB 69 LADLRKYVQAASRRARAGAPWARKKHGIFGVISAPEVFFSSNGMHHIHALLRGT 128
 QY 181 VSENILESFSDAMFDRMTSKLVISGFAPLRNSGGLDVRRKIGEBADQVLAAYITKTASGV 240
 DB 129 GAE--AODLGEWFVRRTLEAIQAAGYAL--DQGVSYI--RSEKLAIFYSK--GV 178
 QY 241 GME-----VSGDGKSGRHGNRAPEIADVAGDPQALMREFEFGSMGRRAIAS 293
 DB 179 GRTIDIAWEMAGQATKARSRNEGHPPEI--LELASGDGKALMLLEYAAMKGVRSCTVS 237
 QY 294 RGLRARAAGLGAELTDQIVOE 315
 DB 238 PAIAAALGIEADEDEERERGERE 259

RESULT 13
 Q52220 PRELIMINARY; PRT; 396 AA.
 ID Q52220

AC Q52220;
 DT 01-NOV-1996 (Tremblrel. 01, Created)
 DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
 DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
 DE Replication protein.
 GN REPA.
 OS Acinetobacter baumannii.
 OG Plasmid pAB49.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
 OC Moraxellaceae; Acinetobacter.
 CX NCBI_Taxid=470;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Fu U.-F.;
 RT "Sequence analysis of an Acinetobacter baumannii cryptic plasmid pAB49."
 RL Submitted (May-1996) to the EMBL/GenBank/DBJ databases.
 DR EMBL; L77992; AAA9423.1; -.
 DR InterPro; IPR000989; Rep.
 DR Pfam; PF01446; Rep; 1.
 KW plasmid.
 SQ SEQUENCE 396 AA; 45656 MW; F08662DD1ADEBA20 CRC64;

Query Match 13.8%; Score 277.5; DB 2; Length 396;
 Best Local Similarity 27.5%; Pred. No. 3.3e-13;
 Matches 74; Conservative 51; Mismatches 109; Indels 35; Gaps 8;

QY 67 FCGLRSCGKMTCPCCAGKVGANRADEISQVVAHQLTGTG--SVANVTMTMTTACQRLHDL 124
 DB 8 YGNVQIGSTWSPVCAKQOTQRRNELGKESWKTVNGSYLLTLTFSSHSPQSLKS 147
 QY 125 LMTGLSAAMKATNGRMRTTEREMYGCDGVYRAVEITTHGKNGHVVHALLMFGSDVSEN 184
 DB 148 NLBGLKRAMRFRYETTRVQAIFPKLSVPHKIKGLVETVQNGMHPHHVLL----- 198
 QY 185 ILSEFSDAMFDRMTSKLVISGFAPLRNS-----GGLDVRRKIGEBADQVLAAYITKTA 237
 DB 199 -LAEHHDLRKDYTSBTLTETLWTKACIKSGUNPMSKRGILDR--GSYADQVYSK----- 251
 QY 238 SGVGMVSGDGKSGRHGNRAPEIADVAGDPQALMREFEFGSMGRRAIA 291
 DB 252 -GLEDELSTKHVKKNGNGFTFDLNFSTIEDNETYGGKPS--KLFGFAISMKGARQLV 308
 QY 292 WSRGLRARAAGLGAELTDQIVOEESAPV 320
 DB 309 WSRGLKRLGI--EKSDEBLAVETDYAS 336

RESULT 14

Q57477 PRELIMINARY; PRT; 528 AA.
 AC Q57477; O09454; O09409;
 DT 01-NOV-1996 (Tremblrel. 01, Created)
 DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
 DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
 DE REP protein.
 GN REP.
 OS Streptomyces phaeochromogenes.
 OG Plasmid pUV1.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Streptomycinae; Streptomycetaceae; Streptomyces.
 CX NCBI_Taxid=1923;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX STRAIN=NRRL-B3559; PubMed=7582009;
 RX MEDLINE=96036207; PubMed=7582009;
 RA Servin-Gonzalez L., Samperi A.I., Cabello J., Galvan L., Suarez V.,
 RA Castro C.;
 RT "Sequence and functional analysis of the Streptomyces phaeochromogenes
 RT plasmid pUV1 reveals a modular organization of Streptomyces plasmids
 RT that replicate by rolling circle";
 RL Microbiology 141:2499-2510(1995).
 RN [2]

RP SEQUENCE FROM N.A.
 RC STRAIN=NRRL-B3559;
 RX MEDLINE=94052369; PubMed=8234485;
 RA Servin-Gonzalez L.;
 RT "Relationship between the replication functions of Streptomyces
 RT plasmids pSV1 and pD101.";
 RL Plasmid 30:131-140(1993).
 RN [3]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92192466; PubMed=1547947;
 RA Denis F., Brzezinski R.;
 RT "A versatile shuttle cosmid vector for use in Escherichia coli and
 RT actinomycetes.";
 RL Gene 111:115-118(1992).
 RN [4]
 RP SEQUENCE FROM N.A.
 RA Denis F., Brzezinski R.;
 RL Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: ESSENTIAL FOR THE AUTONOMOUS REPLICATION OF THE PLASMID
 CC PTV1.
 DR EMBL: U23762; AAA91008.1; -
 DR EMBL: AF004429; AAB60866.1; -
 DR InterPro: IPR000989; Rep.
 DR Pfam: PF01446; Rep. 1.
 KW Plasmid; Plasmid partition.
 SQ SEQUENCE 528 AA; 58079 MW; DBA3464D140EDB2F CRC64;
 Query Match 13.3%; Score 266.5; DB 2; Length 528;
 Best Local Similarity 24.8%; Pred. No. 3.4e-12;
 Matches 100; Conservative 45; Mismatches 137; Indels 121; Gaps 17;
 QY 34 LQOITTSYFNACGR---PISGVNGVTIVNGPKSGSGFG---GIRSCGKGMICPCGAKV 86
 DB 2 LNRVSGIDACGCGCGRVLDPDPTGV---IYAKSSRGVVTIGVRCGRIMFCPCSSAI 56
 QY 87 GAHRADRI-SQVVAHQLTGSAVAMTMTMHTAGORLHD---LWNG-----128
 DB 57 RRGRTBELITGALRLAAGTILAVVLTARNTQTTDLSVLAALMGPLDDKGAFLVDR 116
 QY 129 -----LSAAMKAAATNGRWRTEREMV-----GDDGYRAVEIT 161
 DB 117 SGKPRRAGAYQRMLTAPAFYGRPRARTRFKDQYVAPARDGIRHRIYIGMVRRAEVT 176
 QY 162 HG-KXNGMNVHVALMFGSDV-----SENILESDDMFPRWTSKL-----201
 DB 177 RSKKNGYHPLMLLVFLGSEUSGTPAKDGVVGHFEPSTTDIGWEDMLREMWAGALKRAD 236
 QY 202 ----VSLGPAAP---LNSG--GLDVRKIGRADQVLAAYLTR-----I 236
 DB 237 PKFEPSTDDTGGCKCKGKGVWYSIVRSADVALLLEYLTKNDGKRERPDVDLRA 296
 QY 237 ASGVGMEVSGSGKSGR-HGNRAPELA-----VDVAGDPQALB-----LMREF 280
 DB 297 AGAAAMETARLDSKTKGRKRSMTPEQILYRLMDIEVAGLDPDMEGSGTPEKQLRAMWAQY 356
 QY 281 EFGSMGRRAIAMSRLRARAGIGA---ELTDAQIVEQESAPV 320
 DB 357 EEALAGRRIEWTGRLRRHVLDLSDDDDETDLQYVPEPEAPRL 399
 RESULT 15
 QBRNS3 PRELIMINARY; PRT; 478 AA.
 AC QBRNS3;
 DT 01-JUN-2002 (Tremblrel. 21, Created)
 DT 01-JUN-2002 (Tremblrel. 21, last sequence update)
 DT 01-JUN-2002 (Tremblrel. 21, last annotation update)
 DE Putative replicase.
 GN REP.
 OS Corynebacterium glutamicum (Brevibacterium flavum).
 OC Plasmid PX2608.
 OC Bacteria; Actinobacteria; Actinobacteriales; Actinomycetales;
 OC Corynebacteriaceae; Corynebacterium.

OK NCBI_TaxID=1718;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=227;
 RA Lei C., Zheng Z., Ren Z.;
 RT "Identification, sequencing and characterization of a new plasmid
 RT PX2608 from Corynebacterium glutamicum 227.";
 RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF479770; AAL8305.1; -
 KW Plasmid.
 SQ SEQUENCE 478 AA; 52351 MW; 644F16C27DDA297D CRC64;
 Query Match 13.2%; Score 265; DB 2; Length 478;
 Best Local Similarity 25.1%; Pred. No. 3.9e-12;
 Matches 98; Conservative 49; Mismatches 143; Indels 100; Gaps 17;
 QY 25 GIRHRLRPKLQOITTSYFNACGR---PISGVNGVTIVNGPKSGFGSLRSCGKGMICPC 82
 DB 104 GLRGEVR-----VCGTVPLVNGVGLSVTTG-NGRVSGMWRGCSRM-CPEC 148
 QY 83 AGVGAHRADRI-SQVVAHQLTG-SVAMTMTMHT-----TAGORLHD-----L 125
 DB 149 RAKVAKAKAAVEBOAVCTALDRDLIVMTTLTASHYTKDLRAAGSLHEAVQVTTQV 208
 QY 126 WTGLSAAMKAAATNGRWRTEREMVGGDGYRAVEIT-----HGKGMNVHVALMF- 177
 DB 209 RTEGRLVALATGRRGADLR--GRVGMITAGVTTTDLIVPASRTGIHMRHILVLE 266
 QY 178 --SGDVSENILESDDMFPRWTSKLVSIGFAAPLNSGGLDVR-KIGRADQVLAAYLT 234
 DB 267 FQBLUTTKQVANEYGERLFEYWOQCEGAGLVA---DKRGFTTATGTESALIELAGIVA 323
 QY 235 KIAS-----GVGMEVSGDGKSGRGNRAPE-----IADVAGDPQALE- 275
 DB 324 KGESPQAKAGEITTLHMAHPEGKKGGRSVSEQVLRNLAICYDGLGDEDKRYAR 383
 QY 276 ---LMREFSGMGRRAIAMSRLRARAGIETLTAQIVEQES-SAPVMAIIPKRSVM 331
 DB 384 LVAQMKLVEGTGKVMHRLRSPGLRDLVGLGSELSDELIANTEDWEAASPVAVV-----437
 QY 332 MIRTCAVYFGEILGLVEAGATWENTRDHL 361
 DB 438 -----KWEDELEHV 446

Search completed: November 8, 2003, 23:19:23
 Job time: 69 secs